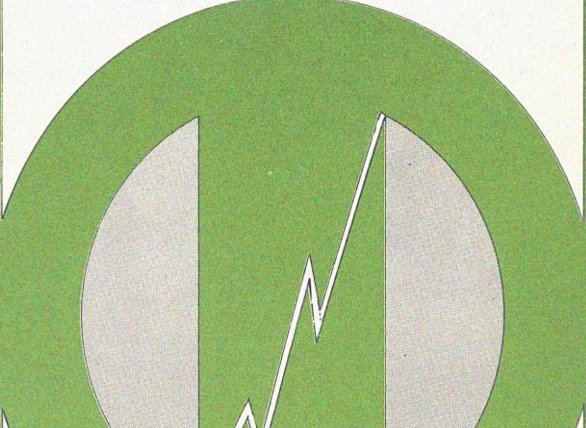
برنار غريه

طرق الإحصاء



ترجمة

هيثم لمع

طـرق الاحصـاء

الطبعة الأولى

1409 هـ - 1989 م

المؤسسة الجامعية للدراسات والنشر والتوزيع

بيروت - الحمراء - شارع اميل اده - بناية سلام
هاتف: ٨٠٢٤٢٨ - ٨٠٢٤٠٧ - ٨٠٢٢٩٦
بيروت - المصيطبة - بناية طاهر - هاتف ٣٠١٠٣٠ - ٣١١٣١٠
ص. ب ٦٣١١ / ١١٣ تلكس LE ٢٠٦٦٥ - ٢٠٦٨٠

الكتاب للنشر والتوزيع

برنار غريه

طرق الإحصاء

ترجمة

هيثم لمع

المؤسسة الجامعية للدراسات والنشر والتوزيع

هذا الكتاب ترجمة

méthodes statistiques

Par

Bernard Grais

© BORDAS, Paris

تمهيد

لقد وُضع هذا الكتاب بهدف سد ثغرة معيّنة . ففي الواقع يوجد العديد من الكتب الممتازة ، إن بالفرنسية أو الإنكليزية ، التي تهتمّ بالإحصاء الوصفي وحساب الاحتمالات والإحصاء الرياضي والتي تناسب مختلف مراحل التعليم التقليدي للإحصاء . من جهة أخرى ، نجد كتباً متخصصة بهذا التطبيق الإحصائي أو ذاك : طريقة الأبحاث الإحصائية (sondages) ، فحص المصنوعات ، فحص المحاسبة ، الخ . إلاّ أنّه لا يوجد ، حسب معرفتنا ، كتاب يقدّم بشكل عملي وموجّه عمداً نحو التطبيقات ، التأليف بين كل هذه المظاهر التي يتمّم أحدها الآخر لنمط التفكير الإحصائي . يضطر إذن الطالب وذو الخبرة اللذان يسعيان لاكتساب ممارسة التقنيات الإحصائية للإطّلاع على سلسلة من الأعمال غالباً ما يختلف مستواها وطرق عرضها ودلالاتها ، وهذا ما يجعل المهمّة صعبة أحياناً .

من ناحية أخرى ، عندما لا يكون مستوى هذه الكتب نموذجياً بشكل يسمح بالتوجه بسهولة نحو التطبيقات العملية ، فإنّها تقدّم عامّة درجة من التشدّد الرياضي تُنفر القارىء دون أن تكون ، معظم الأحيان ، ضرورية فعلاً لفهم الفكرة المطروحة ولتنفيذ التطبيقات .

يطمح هذا الكتاب إذن أن يعطي ، تحت صورة عملية ودون رجوع مبالغ فيه إلى الأداة الرياضية ، عرضاً متكاملاً للتقنيات الإحصائية الضرورية اليوم للمسؤولين والكوادر في الأعمال المختلفة .

لقد كان الكتاب الأوّل « الإحصاء الوصفي » مكرّساً للطرق النموذجية ،

الوصفية بشكل خاص ، التي تكفي غالباً لتأويل المعطيات المتوفرة لتوضيح وتسهيل أخذ القرارات .

هذا الكتاب الثاني يقدّم أدوات التحليل التي يجب اللجوء إليها في حالات أكثر تعقيداً . تعتمد هذه المناهج أو الطرق بغالبيتها على حساب الاحتمالات . من هنا كانت الاستعانة بالمبادىء الرياضية أهم منها في الكتاب الأول الإحصاء الوصفي .

إلاّ أنّنا اعتمدنا أقلّ كمّية ممكنة من التوسعات الرياضية ، وهي كمية ضرورية لعرض متين للمفاهيم ولتبرير النتائج . وبإمكان القارىء الذي يهتمّ بشكل خاص بالمبادىء والنتائج والتطبيقات أن يهملها دون مشكلة .

إضافة إلى ذلك، فإنّ تطوّر الصعاب مدرّج بعناية، كما عالجنا الأمثلة، التي أوردناها كثيرة ، باهتمام خاص وعرضناها بطريقة موسّعة بغية إعطاء القارىء غير المتآلف مع الطرح الرياضي ، تمثيلاً محسوساً لأفكار الكاتب ودليلاً للتطبيق على حالات من الواقع .

إسمحوا لي أخيراً أن أقدّم شكري مجدّداً إلى كلّ الذين ساهموا بتحقيق هذا العمل : السيد ريمون دوما ، المدير العام السابق للمكتب الإحصائي لدول السوق الأوروبية الذي سهّل مهمتي بدرجة كبيرة وأغنى طروحاتي بإتاحته لي استعمال كتابه « الأعمال والإحصاء » كنقطة انطلاق ؛ السيد أندريه ـ برونيه ، الأستاذ في المعهد الوطني للفنون والمهن الذي شجّعني في مهمّتي وأفادني بنصائحه ؛ السيدة مونيك باسّاجيه والآنسة آنيك ميرليه اللتان أخذتا على عاتقهما أمر تقويم المخطوطة وشاركتا بإعادة قراءة التجارب ؛ أخيراً كلّ زملائي الذين أمدّوني بمعلوماتهم القيّمة حول هذه النقطة أو تلك . أتمنّى أن يجد الجميع هنا عبارة عرفاني بالجميل الخالصة .

ب. غريه

الفصل الأول

مدخل إلى حساب الاحتمالات

لقد عرضنا في الكتاب الأوّل الإحصاء الوصفي الطرق الكفيلة بترتيب الملاحظات الإحصائية حسب توزيعات معيّنة وتمثيلها بيانياً وإيجازها من خلال ميّزات ذات ميل مركزي وميزات تفرّق (dispersion) أو من خلال الدلائل الإحصائية في حالات السلاسل المعقدة . ولا يجب إساءة تقدير فعالية هذه الطرق الوصفية البحتة : فهي تسمح بإجراء التقريبات والمقارنات وتسلّط الضوء على خاصيات مهمّة لولاها قد تبقى طي الكتمان . في معظم الأحيان ، تكفي هذه التقنيات النموذجية لتسهيل أخذ القرارات خلال مهمّة ما .

يبقى أن نجتاز خطوة مهمّة : وهي ، في حالات معيّنة ، تمثيل الظواهر الملحوظة بواسطة نماذج تعتمد على الاحتمالات ، أي بواسطة « قوانين إحصائية » تسمح بحساب احتمال حدث معيّن . فهكذا نستطيع حلّ نوع جديد من المعضلات : التقديرات (estimations) والفحوص التي نجريها على عيّنة (échantillon) ما (فحص نوعية إنتاج معيّن أو دقّة حسابات معيّنة) وتنظيم إنتاج البضائع ، الخ . .

إنّ تحديد هذه « القوانين النظرية » يستند إلى مفهوم الاحتمال .

لهذا قبل أن نشرع بدراسة جدول القوانين الرئيسية المعتمدة لشرح الظواهر الإحصائية ، سنكرّس هذا الفصل لمقدّمة نموذجية عن حساب الاحتمالات . في أيامنا هذه ، يُقدَّم حساب الاحتمالات انطلاقاً من نظرية مبدئية تعتمد بدرجة واسعة على لغة المجموعات . وكي نبقى مخلصين لمبدأ الكتاب ، فضّلنا أن نبقى قريبين من الواقع الملموس وأن نقدّم مفهوم الاحتمال إنطلاقاً من أمثلة بسيطة استعرناها من ألعاب الصدفة ومن خلال اعتمادنا على مفهوم الحوادث النموذجية متعادلة الاحتمال .

القسم I : المفهوم البديهي للاحتمال

تاريخياً ، انبثق مفهوم الاحتمال عن أمثلة بسيطة مستعارة عامة من الألعاب التي تعتمد على الصدفة .

1 ـ إذا رمينا قطعة نقود في الهواء ،فإنّ هذه العملية تمثّل اختباراً، أي تجربة لسنا أكيدين من نتيجتها . هناك إمكانيتان : الوجه pile أو الوجه face .

إذا كانت القطعة متناسبة الشكل ومرميّة فعلاً بلا قصد معيّن ، بإمكاننا التصوّر أنّ هاتين الإمكانيتين هما متعادلتا الاحتمال .

لنأخذ إمكانية « الحصول على الوجه face » . بين النتيجتين متعادلتي الاحتمال لا تناسبنا سوى واحدة وهي الحصول على الوجه face . إذن احتمال الحصول على الوجه face يساوي 1/2 .

2 ـ إذا أردنا سحب ورقة من ورق اللعب الذي يتألّف من 52 ورقة ، فإنّنا لا نستطيع مسبقاً معرفة الورقة التي ستُسحب . إذا كان الورق مخلوطاً جيّداً والسحب بلا قصد معيّن ، فإنّ كلّ الأوراق لها نفس الحظ بأن تُسحب : هناك 52 إمكانية متعادلة الاحتمال ، واحتمال الحصول على ورقة معيّنة ، أس الكبّة مثلاً ، يساوي 1/52 .

3 ـ بشكل عام أكثر ، في حال وجود n إمكانية تتنافى إحداها مع الأخرى ومتعادلة الاحتمال جميعها نتيجة اختبار ما (رمي قطعة نقود ، سحب ورقة لعب ، الخ) . وإذا كان بينها k إمكانية مؤاتية (مناسبة) لحدث A معيّن (مثلاً ، سحب ورقة كبة) ، فإنّ احتمال هذا الحدث يساوي k/n :

$$P(A) = \frac{k}{n} = \frac{\text{عدد الإمكانيات المناسبة المتعادلة الإحتمال}}{\text{عدد الإمكانيات المحتملة المتعادلة الإحتمال}}$$

تُسمى الإمكانيات أيضاً أحداثاً نموذجية وتؤلّف مجموعة كلّ الإمكانيات المحتملة مجموعة الأحداث .

أمثلة

ـ لنسحب ورقة من ورق لعب يتألّف من 52 ورقة . ما هو احتمال أن نسحب ورقة كبّة ؟

$$P(\text{سحب ورقة كبّة}) = \frac{13}{52} = \frac{1}{4}$$

يوجد في الحقيقة 13 ورقة كبّة في الورق . هناك إذن بين الإمكانيات الـ 52

المحتملة والمتعادلة الاحتمال 13 إمكانية مناسبة للحدث الذي نريد .

ما هو احتمال أن نسحب ملكاً ؟

$$p\{\text{سحب ملك}\} = \frac{4}{52} = \frac{1}{13}.$$

- إذا رمينا حجر زهر ، ما هو احتمال أن نحصل على نقطة مفردة ؟

$$p\{\text{نقطة مفردة}\} = \frac{3}{6} = \frac{1}{2}.$$

بين الإمكانيات الست المحتملة والمتعادلة الاحتمال ، يوجد في الحقيقة ثلاث (الواحد ، الثلاثة والخمسة) تناسب الحصول على نقطة مفردة .

- وضعنا في وعاء 10 كرات بيضاء ، 20 كرة سوداء و30 كرة حمراء لا يمكن التميز بينها جميعاً بواسطة اللمس وموضوعة بلا ترتيب معيّن . نسحب كرة واحدة :

$$p\{\text{سحب كرة بيضاء}\} = \frac{10}{60} = \frac{1}{6}$$

$$p\{\text{سحب كرة سوداء}\} = \frac{20}{60} = \frac{1}{3}$$

$$p\{\text{سحب كرة حمراء}\} = \frac{30}{60} = \frac{1}{2}$$

الإستحالة . التأكّد

لنفترض أنّه أُعلن عن سحب تومبولا يتألّف من 1000 بطاقة ، نسحب منها واحدة رابحة .

إنّ احتمال أن يربح الجائزة شخص لم يشترِ أيّ بطاقة يساوي انطلاقاً من تحديدنا $\frac{0}{1000} = 0$. إذن فاحتمال حدث مستحيل هو صفر . إنّه الاحتمال المنسوب إلى الجزء « الفارغ » من مجموعة الأحداث . لنفترض أن شخصاً قد اشترى جميع البطاقات ، احتماله لأن يربح الجائزة يساوي $\frac{1000}{1000} = 1$. احتمال الحدث الأكيد يساوي إذن 1 . إنّه الاحتمال المنسوب إلى مجموعة الأحداث نفسها . وبين هاتين الحالتين القصويين يوجد سلّم باقي الأحداث المحتملة .

الإحتمال هو إذاً دائماً محصور بين 0 و1 .

$$0 \leqslant P \leqslant 1.$$

ملاحظة : إنّ مجموع احتمالات جميع الأحداث الممكنة والمتنافية في ما بينها يساوي 1 .

لنعد إلى مثل الوعاء حيث يمكننا أن نسحب كرة بيضاء أو سوداء أو حمراء وليس هناك أية إمكانية أخرى . نرى جيّداً أنّه :

p { كرة بيضاء } + p { كرة سوداء } + p {كرة حمراء} = 1

$$1^{(1)} = \frac{1}{2} + \frac{1}{3} + \frac{1}{6}$$

إنه احتمال مجموعة الأحداث .

الحدث المتمّم

يتألّف الحدث المتمّم لحدث A معيّن من جميع الإمكانيات المحتملة والمتنافية والتي لا تشكّل جزءاً من A . إنّه متمّم A في مجموعة الأحداث . لنأخذ ، في المثل السابق ، احتمال أن نسحب كرة سوداء أو كرة حمراء .

بإمكاننا التفكير مباشرة بهذه الطريقة :

p { سوداء أو حمراء } = عدد الحالات المناسبة / عدد الحالات المحتملة $= \frac{20 + 30}{60} = \frac{5}{6}$.

ولكن يمكننا اعتماد طريقة التفكير التالية :

الحدث المتمّم هو: سحب كرة بيضاء . في الواقع إنّ هاتين الإمكانيتين : « سحب كرة بيضاء » و« سحب كرة سوداء أو حمراء » تغطّيان كامل حقل المحتمل .

إذن : p { بيضاء } + p {سوداء أوحمراء} = 1

p { سوداء أو حمراء } = 1 − p { بيضاء }

$$= 1 - \frac{10}{60} = \frac{5}{6}$$

في بعض الأحيان ، قد يكون احتمال الحدث المتمّم أسهل للحساب، من هنا أهمّية هذه الطريقة .

نستنتج إذن أنّه في الحالات العادية ، حساب الاحتمال هو عبارة عن حساب عدد الحالات المحتملة المتعادلة وحساب عدد الحالات المناسبة لتحقيق حدث معيّن .

مثلاً : نسحب 13 ورقة من ورق لعب مؤلّف من 52 ورقة . ما هو احتمال سحب كلّ أوراق الكبّة ؟

للإجابة عن هذا السؤال ، يجب أن يكون بإمكاننا أن نحسب عدد الإمكانيات

(1) يُرجى قراءة المعادلات والعبارات والمباينات والجداول المكتوبة باللاتينية ، على مرّ الكتاب ، من اليسار إلى اليمين .

المحتملة ومتعادلة الاحتمال التي يتضمّنها سحب 13 ورقة بين 52 . وهذا ما يقودنا إلى دراسة معضلات التعداد أي التحليل التوافيقي (analyse combinatoire) .

القسم II : فكرة عامّة عن التحليل التوافيقي

1 . التبديلات ـ 2 . الترتيبات ـ 3 . التوافقيات

يهدف التحليل التوافيقي إلى تعداد مختلف التشكيلات التي نستطيع إجراؤها إنطلاقاً من مجموعة عناصر . وهو يسمح لنا بحساب عدد الإمكانيات متعادلة الاحتمال المرتبطة باحتمال معيّن ، مثلاً سحب 13 ورقة لعب بين 52 ورقة . في ما يلي ، سنرمز إلى العناصر بواسطة حروف أبجدية .

التشكيلات المرتّبة وغير المرتّبة

- التشكيلات المرتّبة : في هذه الحالة نعتبر أن تشكيلين يتألّفان من نفس العناصر هما مختلفان إذا لم تحتلّ هذه العناصر نفس الأمكنة في كلّ منهما .

مثلاً . التشكيلان (a, b) و(b, a) هما مختلفان إذا أخذناهما كتشكيلين مرتّبين .

- بالمقابل فإن تشكيلين غير مرتّبين يُعتبران واحداً في حال تألّفا من نفس العناصر.

مثلاً : التشكيلان (a, b) و(b,a) هما نفسهما إذا أخذناهما كتشكيلين غير مرتّبين .

سندرس في ما يلي أنواعاً ثلاثة من التشكيلات : التبديلات ، الترتيبات والتوافقيات .

I . التبديلات (Permutations)

إذا أخذنا العناصر الثلاثة a ، b وc ، بإمكاننا إجراء التبديلات التالية :

$$\left.\begin{matrix} abc & bac \\ acb & bca \\ cab & cba \end{matrix}\right\} \text{6 تبديلات}$$

التبديل هو تشكيل مرتّب لأنّ كلّ تبديل يتضمّن كل العناصر لا يتميّز إلاّ بالمكان اذي تأخذه هذه العناصر .

تعريف . التبديل الذي يتألّف من n عنصراً هو تشكيل مرتّب لمجموعة هذه العناصر حيث يظهر كلّ منها مرّة واحدة فقط .

لنسمِّ p_n عدد التبديلات الممكن إجراؤها بواسطة n عنصراً :

$$\boxed{P_n = 1 \times 2 \times 3 \times \cdots \times n = n!}$$

(إقرأ : p_n يساوي عاملية n (factorielle n) . وتساوي « عاملية n » التي نرمز إليها بـ n! حاصل ضرب الـ n عدداً الصحيحة الأولى) .

البرهان : في حال عنصر واحد :

a

من خلال عنصر واحد يمكننا إجراء تبديل واحد .

في حال عنصرين : بإمكاننا أن نضع العنصر الإضافي على يمين أو يسار العنصر الأوّل ، أي بطريقتين مختلفتين :



إذن نجد من خلال عنصرين تبديلين اثنين .

ثلاثة عناصر : في كلّ من التبديلين السابقين بإمكاننا وضع العنصر الإضافي الثالث بثلاث طرق مختلفة :



من خلال ثلاثة عناصر نجد إذن : 3! = 2 × 3 تبديلاً .

....

n عنصراً :

في كل من الـ p_{n-1} تبديلاً السابقة والتي أُجريت على (n − 1) عنصراً ، بإمكاننا وضع العنصر رقم n في n مكاناً ممكناً :



$$P_n = nP_{n-1}$$
$$P_{n-1} = (n-1)P_{n-2}$$
$$\cdots\cdots$$
$$P_2 = 2P_1$$
$$P_1 = 1$$

إذن :

$$P_n = n(n-1)\ldots 2.1 = n!.$$

هكذا ، فإنّ n عنصراً تعطينا n! تبديلاً .

مثلاً : قطار يتألّف من 10 عربات ، بكم طريقة يمكن تركيب هذا القطار (نفترض أن القاطرة تبقى دائماً في المقدّمة) ؟

$$10! = 3\,628\,800$$

2 . الترتيبات (Arrangements)

لنأخذ العناصر الأربعة a, b, c و d ، ولنرتّبها اثنين اثنين :

$$\left.\begin{matrix} ab & ba \\ ac & ca \\ ad & da \\ bc & cb \\ bd & db \\ cd & dc \end{matrix}\right\} \text{12 ترتيباً}$$

تعريف : إنّ ترتيب p عنصراً اخترناه من بين n عنصراً هو تشكيل مرتّب لـ p من n عنصراً ، حيث كلّ واحد منها يظهر مرّة واحدة على الأكثر في نفس الترتيب .

إذا رمزنا بـ A_n^p إلى عدد ترتيبات p عنصراً مختاراً من بين n :

$$A_n^p = \frac{n!}{(n-p)!}$$

البرهان . إذا أخذنا n عنصراً ، فإنّنا نستطيع معها إجراء ترتيبات تتألّف من 1 2 ..... أو n عنصراً .

الترتيبات بعنصر واحد هي :

$$a, b, c, \ldots, n .$$

$$A_n^1 = n .$$

يمكننا بواسطة n عنصراً إجراء n ترتيباً يتألّف كلّ منها من عنصر واحد .

الترتيبات بعنصرين :

نحصل عليها بوضعنا إلى يمين كل من العناصر السابقة ، a مثلاً ، واحداً من الـ (n−1) عنصراً الباقي :

$$\underbrace{ab, ac, \ldots, an}_{n-1}$$

بالتالي :

$$A_n^2 = (n-1)\,A_n^1$$

يمكننا بواسطة n عنصراً إجراء n (n−1) ترتيباً يتألّف كلّ منها من عنصرين اثنين .

. . . .

ترتيبات بـ p عنصراً

ونحصل عليها بوضعنا إلى يمين كلّ من الـ A_n^{p-1} ترتيباً السابقة والتي يتألّف كلّ منها من (p−1) عنصراً ، واحداً من الـ (p−1) −n عنصراً غير المستعملة .

بالتالي :

$$A_n^p = (n - p + 1)\, A_n^{p-1},$$

ونستنتج من هذا ، بالتكرار:

$$\begin{aligned}
A_n^p &= (n - p + 1)\, A_n^{p-1}\\
A_n^{p-1} &= (n - p + 2)\, A_n^{p-2}\\
&\cdots\cdots\\
A_n^2 &= (n - 1)\, A_n^1\\
A_n^1 &= n\\
\hline
A_n^p &= (n - p + 1) \times \cdots \times (n - 1) \times n\\
&= \frac{n\,!}{(n - p)\,!},
\end{aligned}$$

وذلك إنطلاقاً من تعريف العامليات .

إذن بإمكاننا بواسطة n عنصراً إجراء $\frac{n\,!}{(n-p)\,!}$ ترتيباً يتألّف كلّ منها من p عنصراً .

مثلاً : تقدّم 12 مرشّحاً لانتخابات مجلس إدارة 8 مراكز . إذا أردنا نشر لائحة أسماء المنتخبين تبعاً لعدد الأصوات الحاصل ، كم يبلغ عدد اللوائح الممكنة ؟ (تلعب طريقة الترتيب دوراً) .

$$A_{12}^8 = \frac{12\,!}{4\,!} = 19\,958\,400$$

3 . التوافقيات (Combinaisons)

لنأخذ العناصر الأربعة c, b, a و d ونركّبها اثنين اثنين :

$$\left.\begin{array}{l} ab\\ ac\\ ad\\ bc\\ bd\\ cd \end{array}\right\} 6 \text{ توافقيات}$$

الأمر هو إذن عبارة عن عملية شبيهة بعملية الترتيب ، ولكن هذه المرّة يُعتبر تشكيلان يتضمّنان نفس الأحرف متشابهين مهما كانت أماكن وجود هذه الأحرف : التوافقية هي تشكيل غير مرتّب .

تعريف : إن توافقية p عنصراً اخترناه من بين n عنصراً هي تشكيل غير مرتّب لهذه العناصر حيث يظهر كلّ واحد منها مرّة على الأكثر .

نرمز بـ C_n^p وأحياناً $\binom{n}{p}$ إلى عدد التوافقيات الممكن إجراؤها بواسطة p عنصراً نختاره بين n .

$$\boxed{C_n^p = \frac{n\,!}{p\,!(n-p)\,!}}$$

البرهان : لنأخذ توافقية p عنصراً نختارها بين n ونرمز إليها بأحرف أبجدية . بما أن التوافقية هي تشكيل غير محكوم بالترتيب ، بإمكاننا كتابته حسب الترتيب الأبجدي :

$$\underbrace{(a, c, f, g, \ldots, k)}_{\text{عنصراً } p}\,.$$

يمكننا انطلاقاً من هذه التوافقية إجراء كل الترتيبات التي تتضمّن الـ p حرفاً (a, b, c, f, g, ..., k) وذلك بتبديلها في ما بينها . يوجد إذن p! ترتيباً من هذا النوع . بإمكاننا إذن ، انطلاقاً من توافقية ما ، إجراء p! ترتيباً . بالتالي :

$$p\,!\,C_n^p = A_n^p = \frac{n\,!}{(n-p)\,!}\,.$$

أي

$$C_n^p = \frac{n\,!}{p\,!(n-p)\,!}\,.$$

هكذا ، يسمح n عنصراً بإجراء $\frac{n\,!}{p\,!(n-p)\,!}$ توافقية يتألّف كلّ منها من p عنصراً .

خصائص التوافقيات

$$1. \qquad C_n^p = C_n^{n-p}\,.$$

وهذا في الواقع ناتج عن تناظر (symétrie) القاعدة :

$$C_n^p = \frac{n\,!}{p\,!(n-p)\,!} = C_n^{n-p}$$

بعبارة أخرى ، بما أنّه لا أهمية لطريقة الترتيب ، فإنّ اختيار p عنصراً بين n هو كاختيار الـ n−p عنصراً التي لا تنتمي إلى التوافقية .

$$2. \qquad C_n^p = C_{n-1}^p + C_{n-1}^{p-1}.$$

لنأخذ n عنصراً : n, ..., b, a

بإمكاننا تأليف كل التوافقيات التي تحتوي العنصر a بإضافتنا إليه (p−1) عنصراً نختاره بين الـ (n−1) عنصراً مختلفاً عن a . إذن يبلغ عدد التوافقيات التي تتضمّن a :

$$C_{n-1}^{p-1}.$$

أمّا عدد التوافقيات التي لا تحتوي a والتي نحصل عليها باختيارنا p عنصراً بين الـ (n−1) عنصراً المختلفة عن a فيبلغ :

$$C_{n-1}^p.$$

بالتالي فإنّ المجموع الكلّي للتوافقيات التي يتألّف كلّ منها من p عنصراً مأخوذاً من n عنصراً هو :

$$C_n^p = C_{n-1}^p + C_{n-1}^{p-1}.$$

تطبيق : مثلّث باسكال

إنّ القاعدة السابقة تعطي طريقة سهلة لحساب قيم C_n^p بالتكرار ، وتُدعى نتيجة هذه الطريقة بمثلّث باسكال (الشكل 1) :

$$C_{n-1}^{p-1} + C_{n-1}^p = C_n^p$$

كلّ عنصر من الجدول هو عبارة عن حاصل جمع العنصر الذي يقع فوقه مباشرة مع العنصر الذي يوجد إلى يسار هذا الأخير .

وكي تملأ علاقة التكرار دورها كلّياً ، وجب علينا أن نتفق على وضع :

$0! = 1$ أي أنّ $C_n^0 = 1$.

ففي الحقيقة

$$C_n^0 + C_n^1 = C_{n+1}^1,$$

$$C_n^0 + n = n + 1,$$

أي

$$C_n^0 = 1.$$

n \ $p \rightarrow$	0	1	2	3	4	5	6 ...
0	1						
1	1	1					
2	1	2	1				
3	1	3	3	1			
4	1	4	6	4	1		
5	1	5	10	10	5	1	
6	1	6	15	20	15	6	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

الشكل 1 ـ مثلث باسكال

3 . عرض ذات الحدّين نيوتن (binôme de Newton) :

$$(p+q)^n = \sum_{k=0}^{n} C_n^k p^k q^{n-k} .$$

البرهان :

$$(p+q)^2 = p^2 + 2pq + q^2$$
$$(p+q)^3 = p^3 + 3p^2q + 3pq^2 + q^3$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$(p+q)^n = p^n + C_n^1 p^{n-1} q + \cdots + C_n^k p^k q^{n-k} + \cdots + q^n$$

في الواقع ، نحصل في هذه العبارة على عنصر يحتوي على $p^k q^{n-k}$ باختيارنا p بين k عاملاً ويؤخذ q بين الـ (n−k) عاملاً الباقية التي تؤلّف $(p+q)^n$ للتمييز بين العوامل ، لنشير إلى كلّ منها بواسطة حرف أبجدي :

$$(p+q)^n = \underbrace{(p+q)}_{\text{العامل a}} \times \underbrace{(p+q)}_{\text{العامل b}} \times \underbrace{(p+q)}_{\text{العامل c}} \times \cdots \times \underbrace{(p+q)}_{\text{العامل n}}$$

بإمكاننا إذن تأليف عدد من العناصر $p^k q^{n-k}$ يبلغ نفس عدد طرق اختيار k عاملاً من a, b, c,... بين n عاملاً . وبما أن طريقة ترتيب العوامل لا تهمّ فإنّنا نحصل على C_n^k عنصر $p^k q^{n-k}$.

ملاحظة : إذا جعلنا في قاعدة ذات الحدين نيوتن :

p = q = 1

فإنّنا نحصل على النتيجة الفريدة التالية :

$$C_n^0 + C_n^1 + \cdots + C_n^n = 2^n$$

إنّ مجموع المعامِلات في ذات الحدين نيوتن يساوي 2^n .

مثل على التوافقيات

تقدّم 12 مرشحاً لانتخابات مجلس إدارة يضمّ 8 مراكز . إذا أردنا نشر لائحة أسماء المنتخبين حسب الترتيب الأبجدي ، كم يبلغ عدد اللوائح الممكنة ؟

$$C_{12}^8 = \frac{12!}{8!\,4!} = 495 .$$

تطبيق التحليل التوافيقي على حساب الاحتمالات

أصبح الآن بوسعنا الإجابة عن السؤال الذي سبق أن طرحناه على أنفسنا : إذا سحبنا 13 ورقة من 52 ورقة لعب ما هو احتمال أن نسحب كلّ أوراق الكبّة ؟

إنّ ورق لعب يتألّف من 52 ورقة يسمح بإجراء C_{52}^{13} توافقية يتألّف كلّ منها من 13 ورقة ، جميعها متعادلة الاحتمال إذا عدلنا في التوزيع ، وورقة واحدة هي المناسبة ، الاحتمال هو إذن :

$$P = \frac{1}{C_{52}^{13}} = \frac{1}{635\,013\,559\,600}$$

القسم III : امتداد لمفهوم الاحتمال

1 . لغة المجموعات : A . تعريفات ؛ B . عمليات منطقية بين أجزاء المجموعة ـ 2 . مبادىء حساب الاحتمالات : A . قاعدة الاحتمالات الكلّية ؛ B . قاعدة الاحتمالات المركّبة ؛ C . الاستقلالية بين حدثين .

لقد انتشر مفهوم الاحتمال انطلاقاً من حالات كان فيها ممكناً ، لاعتبارات تتعلّق بالتناظر(symétrie) ، تحديد مجموعة من الأحداث المتعادلة الاحتمال . وقد وضع باسكال وفيرما ، بشكل خاص ، تصوّراتهما حول حساب الاحتمالات على أساس معضلات ألعاب الصدفة التي طرحها عليهما لاعب ذكي وفضول يُدعى Le Chevalier de Méré . ولكن تدريجياً ، سرعان ما دعت الحاجة إلى توسيع ميدان حساب الاحتمالات إلى معضلات أكثر تعقيداً : ففي مادّة العلوم الاجتماعية والاقتصادية ليس من الممكن عامة تحديد مجموعة من الأحداث المتعادلة الاحتمال . وقد تمّ هذا الامتداد لمفهوم الاحتمال انطلاقاً من نظرية مبدئية : إنّ الاحتمال المنسوب إلى حدث معيّن هو عدد يجب أن يخضع لعدد من الشروط الضرورية أو المبادىء .

وقبل أن نتابع على هذا الأساس دراسة حساب الإحتمالات ، من الضروري أن نلمّ بفكرة عن لغة المجموعات .

1 . لغة المجموعات

A . تعريفات

المجموعة هي جملة من الأغراض أو الأحداث نسمّيها عناصر وتتميّز جميعها بانتمائها إلى هذه المجموعة . ولا يعود يُنظَر إلى عناصر مجموعة ما إلاّ من زاوية إنتمائها إلى هذه المجموعة .

أمّا تحديد المجموعة فيتمّ :

- إمّا عن طريق تعداد عناصرها ، إذا كان عددها متهياً :

مثلاً : المجموعة

$$E = \{ 3, 13, 0, 7, 8 \}$$

هي المجموعة المؤلّفة من العناصر الخمسة المعدودة ،

- إمّا عن طريق بيان خاصيّة مشتركة لكلّ العناصر :

مثلاً : مجموعة الفرنسيين . ينتمي إلى هذه المجموعة كلّ الأشخاص الذين يحملون الجنسية الفرنسية ؛

- إمّا عن طريق إعطاء قاعدة لبناء عناصر المجموعة :

مثل 1 . مجموعة الأعداد الصحيحة

$$N = \{ 0, 1, 2, 3, \ldots \}$$

يحدّد كلّ عدد ينتمي إلى المجموعة N إنطلاقاً من سابقه بإضافة واحد إلى هذا الأخير ؛

مثل 2 . مجموعة التركيبات التي بوسعنا إجراؤها بواسطة 5 أغراض : e, d, c, b, a . تتضمّن هذه المجموعة 32 عنصراً :

$$C_5^0 + C_5^1 + \cdots + C_5^5 = 2^5 = 32$$

الانتماء

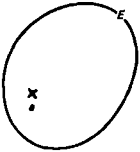
لنفترض e عنصراً من المجموعة E ، عندها نكتب :

$$e \in E ,$$

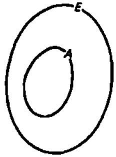
ونقرأ : « العنصر e ينتمي إلى المجموعة E » .

بشكل عام ، نمثّل المجموعة بواسطة مسطّح (مخطّط Venn ، الشكل 2) .

وتُمثّل العناصر بواسطة نقاط داخل هذا المسطّح .



الشكل 2 . مخطّط Venn



الشكل 3

الاحتواء

نقول أنّ المجموعة A محتواة داخل المجموعة E إذا كان كلّ عنصر من A ينتمي أيضاً إلى E (الشكل 3) :

$$e \in A \Rightarrow e \in E .$$

يُقرأ الرمز $\Rightarrow$: « يعني » : « e ينتمي إلى A يعني أنّ e ينتمي إلى E » .

ونكتب عندها

$$A \subset E$$ (« A محتواة داخل E ») .

او :

$E > A$ (« E يحتوي A ») .

ونقول أنّ A هي جزء من E .

ومن خلال تحديد مفهوم الاحتواء نرى أنّ المجموعة E نفسها هي جزء من E . ففي الواقع ، العبارة

$$e \in E \Rightarrow e \in E$$

هي دائماً صحيحة .

المجموعة الفارغة

المجموعة الفارغة هي المجموعة التي لا تتضمّن أيّ عنصر ، ونشير إليها بالرمز $\varnothing$. وقد اتُّفق أنّ المجموعة الفارغة $\varnothing$ هي جزء من E :

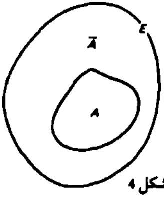
$$\varnothing \subset E$$

مثلاً : إنّ مجموعة التركيبات التي بإمكاننا إيجادها دون اختيار أي غرض بين 5

أغراض هي مجموعة فارغة . إنّها جزء من مجموعة التركيبات التي يمكن الحصول عليها بواسطة 5 أغراض .

المجموعة المتمّمة

لنفترض أن A هي جزء من E ، إن متمّم A بالنسبة للمجموعة E والذي نرمز إليه بـ $\bar{A}$ ، هو مؤلّف من كلّ عناصر E التي لا تنتمي إلى A (الشكل 4) .



الشكل 4

$$e \in \bar{A} \Leftrightarrow e \notin A .$$

الرمز $\Leftrightarrow$ يُقرأ « ما يُعادل » .

مجموعة أجزاء المجموعة

لنأخذ المجموعة التالية :

$$E = \{ a, b, c, d \}$$

ولنكوّن كلّ أجزاء E الممكنة :

$\emptyset$
$\{a\}, \{b\}, \{c\}, \{d\},$
$\{ab\}, \{ac\}, \{ad\}, \{bc\}, \{bd\}, \{cd\},$
$\{abc\}, \{abd\}, \{acd\}, \{bcd\},$
$\{abcd\}.$

هذه الأجزاء تشكّل مجموعة جديدة تُدعى مجموعة أجزاء E ونشير إليها بـ $\mathcal{P}(E)$.

حول هذا الأمر ، لنشر من جديد إلى أنّ المجموعة E نفسها والمجموعة الفارغة $\emptyset$ تنتميان إلى مجموعة أجزاء E :

$$E \in \mathcal{P}(E) \qquad \emptyset \in \mathcal{P}(E) .$$

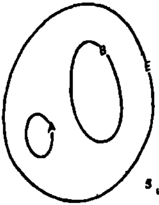
أثناء بحثنا عن أجزاء E ، لاحظنا أنّها مؤلّفة من كلّ التركيبات الممكن إجراؤها بواسطة العناصر المنتمية إلى هذه المجموعة . إذن يبلغ عدد أجزاء مجموعة تتألّف من n عنصراً : 2^n جزءاً .

$C_n^0 + C_n^1 + \cdots + C_n^n = 2^n$ (أنظر القسم II ، الفقرة 3) .

المجموعات المنفصلة

نعتبر أن جزءين A وB من (E)\mathcal{P} هما منفصلان إذا لم يكن بينهما أي عنصر مشترك (الشكل 5) .

إذا كانت عناصر المجموعة E عبارة عن إمكانيات ، فإنّ المجموعات المنفصلة هي أحداث متنافية .



الشكل 5

B . عمليات منطقية بين أجزاء المجموعة

لنفترض أن A وB هما جزءان من نفس المجموعة المرجع E .

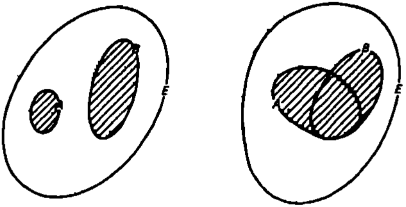
الإتّحاد

الإتحاد بين مجموعتين A وB هو المجموعة R المكوّنة من العناصر المنتمية إمّا إلى A ، إمّا إلى B (بما فيها العناصر التي قد تكون منتمية في الوقت نفسه إلى A وإلى B) (الشكل 6) .

وندلّ إلى الإتحاد بالرمز $\cup$:

$$R = A \cup B .$$

في حال كانت عناصر المجموعة E عبارة عن إمكانيات ، فإن اتحاد جزءين في هذه المجموعة يعني : يتحقّق الحدث $R = A \cup B$ منذ أن يتحقّق على الأقل واحد من الحدثين A أو B .



اتحاد مجموعتين منفصلتين　　　اتحاد مجموعتين غير منفصلتين

المجموعة $R = A \cup B$ هي المخطّطة

الشكل 6

التقاطع

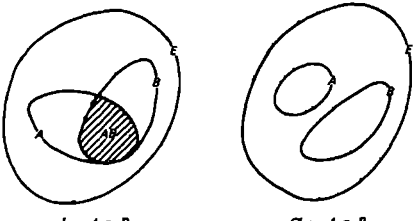
التقاطع بين A وB هو المجموعة I المكوّنة من العناصر التي تنتمي في الوقت نفسه إلى A وإلى B (الشكل 7) .

ندلّ إلى التقاطع بالرمز ∩ :

$$I = A \cap B .$$

إذا كانت المجموعتان A وB منفصلتين ، فإن تقاطعهما يساوي المجموعة الفارغة

$\varnothing$



$I = A \cap B$ $\varnothing = A \cap B$

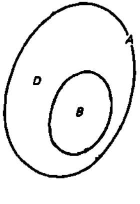
الشكل 7

إذا كانت عناصر المجموعة E عبارة عن إمكانيات فإنّ تقاطع اثنين من أجزاء هذه المجموعة يعني : يتحقّق الحدث $I = A . B$ في حال تحقّق الحدثان A وB على السواء .

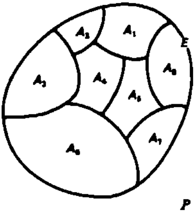
ملاحظة : لنفترض أن A تحتوي B (الشكل 8) ، عندئذٍ :

$$A \cup B = A , \qquad A \cap B = B$$

في هذه الحالة ـ فقط ـ يمكننا تعريف الفارق D = A−B كمجموعة العناصر التي تنتمي إلى A دون أن تنتمي إلى B : $D = A - B \Leftrightarrow A = B + D$



الشكل 8



الشكل 9 : تجزئة المجموعة E
$P = \{A_1, A_2, \ldots, A_8\}$

تجزئة المجموعة

التجزئة P للمجموعة E هي مجموعة الأجزاء

$A_1, A_2, A_3, \ldots, A_k$

غير الفارغة ، المنفصل أحدها عن الآخر والتي يساوي اتحادها المجموعة E (الشكل 9) .

الأجزاء Ai تُدعى فئات التجزئة P .

إنّ عملية تجزئة مجموعة معيّنة تعادل عملية تصنيف أفراد جمهرة (population) ما تحت أسماء معيّنة للفئات أو حسب فئات القيم الممكنة لمتغيّرة إحصائية : كلّ فرد ينتمي إلى فئة واحدة فقط .

في لغة الأحداث ، التجزئة هي تفكيك مجموعة الأحداث إلى أحداث يتنافى واحدها مع الآخر .

التخصيص من مجموعة إلى أخرى

لنأخذ مجموعتين E وF . إنّ المطابقة التي تعطي لكلّ عنصر x من E عنصراً y من F تُدعى تخصيصاً (أو تطبيقاً) من E إلى F (الشكل 10 : أعطينا العنصر 2 من F إلى العنصر a من E ، العنصر 1 إلى العنصر b ، الخ) . قد لا يجد بعض العناصر من F أي مطابق له من E ، كما قد يكون لعنصرين أو أكثر من E المطابق نفسه من F .

نرمز إلى هذا التخصيص بالحرف f ونكتب :

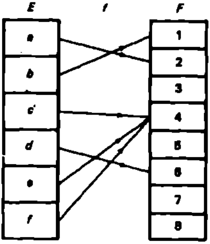
$x \xrightarrow{f} y$ أو $y = f(x)$.

ونقول أنّ y هي صورة x بواسطة f .

لنفترض أن A هي جزء من E :

$A \subset E$.

صورة A هي مجموعة العناصر من F التي تمثّل صوراً لعنصر على الأقل من A .



الشكل 10 : تخصيص من المجموعة
$E = \{ a, b, c, d, e, f \}$
إلى المجموعة $F = \{ 1, 2, \ldots, 8 \}$.

مثلاً : صورة المجموعة $A = \{ c, d, e \}$ بواسطة f هي المجموعة {4, 6} (الشكل 10) .

الصورة المعكوسة

y هو عنصر من F . قد يكون y صورة لعلّة عناصر من E . إنّ مجموعة العناصر x التي تنتمي إلى E والتي تملك y كصورة لها جميعاً تُدعى الصورة المعكوسة للعنصر y ونرمز إليها بـ $f^{-1}(y)$.

مثلاً : المجموعة {c, e, f} هي الصورة المعكوسة لِـ {4} بالنسبة للتخصيص الممثّل في الشكل 10 .

بشكل عام أكثر ، إذا كان H جزءاً من F ، فإنّ الصورة المعكوسة لِـ H هي مجموعة عناصر E التي تنتمي صورها ، بواسطة f ، إلى H .

مثلاً : الصورة المعكوسة للمجموعة {1, 4, 6} هي { b, c, d, e, f } .

2 ـ مبادىء حساب الاحتمالات

إنّ امتداد مفهوم الاحتمال إلى الحالة حيث لا يمكن تحديد مجموعة أحداث متعادلة الاحتمال ، ولكن حيث مجموعة الأحداث متناهية ، لا يمثّل درجة كبيرة من الصعوبة . يكفي في الواقع أن نضع ، كتحديد مبدئي للاحتمال ، القواعد الثلاث التالية ، التي تحفظ لنا الخصائص التي وجدناها سابقاً أي عندما كان باستطاعتنا تعداد الأحداث المتعادلة الاحتمال :

E هي مجموعة متناهية من الأحداث .

1 ـ الاحتمال المنسوب إلى كل حدث (أي إلى كل جزء من E) هو عدد إيجابي أو صفر .

2 ـ الاحتمال المنسوب إلى مجموعة الأحداث E يساوي واحداً :

$$P\{E\} = 1$$

3 ـ لكل زوج (A, B) من الأحداث المتنافية (غير المتوافقة) ، احتمال اتحاد هذين الحدثين يساوي حاصل جمع احتمالي A وB :

$$P\{A \cup B\} = P\{A\} + P\{B\}.$$

وتنتج عن هذه المبادىء قواعد حساب الاحتمالات التي تسمح بإيجاد احتمال حدث معين بواسطة عمليات منطقية نجريها بين أحداث نعرف احتمال كل منها .

A . قاعدة الاحتمالات الكلّية

إنّ قاعدة الاحتمالات الكلّية تعطينا قاعدة حساب احتمال تحقيق واحد على الأقل من حدثين .

حالة حدثين متنافيين

في الحالة حيث الحدثان A وB متنافيان ، أي حيث المجموعتان A وB منفصلتان ، فإنّ قاعدة الاحتمالات الكلّية هي ما رأيناه في المبدأ 3 .

احتمال تحقيق واحد على الأقل من حدثين متنافيين A وB يساوي حاصل جمع احتمالي هذين الحدثين :

$$\boxed{P\{A \cup B\} = P\{A\} + P\{B\}}$$

وتتحقّق ميزة هذا المبدأ بسهولة عندما نستطيع منذ البدء تحديد مجموعة أحداث متعادلة الاحتمال .

لنفترض أنّ A وB هما حدثان متنافيان يُنسب إليهما N_A وN_B حدثاً.تنتمي إلـى مجموعة تتألّف من N حدثاً متعادلة الاحتمال . يُنسب إلى الحدث (A أو B) الذي نرمز إليه بـ $A \cup B$ ، N_A+N_B حدثاً متعادلة الاحتمال ، إذن :

$$P\{A \cup B\} = \frac{N_A + N_B}{N} = \frac{N_A}{N} + \frac{N_B}{N} = P\{A\} + P\{B\}.$$

مثلاً : إذا أردنا سحب ورقة واحدة من ورق لعب يتألّف من 52 ورقة ، ما هو إحتمال سحب بنت أو ملك :

$$P\{\text{بنت أو ملك}\} = P\{\text{بنت}\} + P\{\text{ملك}\}$$

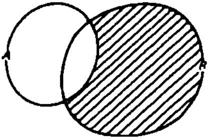
$$\frac{2}{13} = \frac{1}{13} + \frac{1}{13}.$$

بشكل عام أكثر، إذا كان A_1 ، A_2 ، . . . ، A_n أحداثاً يتنافى أحدها مع الآخر، فإنّ مبدأ الاحتمالات الكلّية هو :

$$\Pr\{A_1 \cup A_2 \cup \cdots \cup A_n\} = \Pr\{A_1\} + \Pr\{A_2\} + \cdots + \Pr\{A_n\}$$

حالة حدثين لا يتنافيان

لنفترض أنّ A وB هما حدثان لا يتنافى واحدهما مع الآخر : إذن المجموعتان المنسوبتان إليهما هما غير منفصلتين (الشكل 11) . ولكن نستطيع الوصول إلى حدثين متنافيين باعتمادنا المجموعتين المنفصلتين التاليتين :



الشكل 11

A

$B - (A \cap B)$ (المجموعة المخطّطة)

يمكن القول أنّ :

$$A \cup B = A \cup [B - (A \cap B)].$$

إذن، إذا طبّقنا قاعدة الاحتمالات الكلّية بالنسبة لمجموعتين منفصلتين (المبدأ 3) :

$$P\{A \cup B\} = P\{A\} + P\{B - (A \cap B)\}.$$

الحدثان $A \cap B$ و $(B - A \cap B)$ هما متنافيان :

$$B = [B - (A \cap B)] \cup (A \cap B),$$

$$P\{B\} = P\{B - (A \cap B)\} + P\{A \cap B\}.$$

إذن :

$$\boxed{P\{A \cup B\} = P\{A\} + P\{B\} - P\{A \cap B\}}$$

مثلاً : إذا سحبنا ورقة من ورق لعب (52 ورقة)، ما هو احتمال أن نحصل على ورقة كبة أو ملك :

$$P\{\text{كبّة أو ملك}\} = P\{\text{كبّة}\} + P\{\text{ملك}\} - P\{\text{ملك الكبّة}\}$$

$$\frac{4}{13} = \frac{1}{4} + \frac{1}{13} - \frac{1}{52}$$

في الحقيقة ، يحتوي احتمال سحب ورقة كبّة على احتمال سحب ملك الكبة ؛ كذلك الأمر بالنسبة لاحتمال سحب ملك . إذن يُحسب احتمال سحب ملك الكبّة مرّتين : يجب تنقيصه مرّة واحدة .

B قاعدة الاحتمالات المركّبة

تعطينا قاعدة الاحتمالات المركّبة قاعدة حساب احتمال تحقيق حدثين في آن واحد . وهي تدفعنا أوّلاً إلى تعريف الاحتمال المشروط لحدث معيّن .

الاحتمال المشروط

تعريف : لنفترض أن E هي مجموعة أحداث حدّد عليها احتمال وB حدث ذو احتمال يختلف عن الصفر .

نسمّي احتمال الحدث A المشروط والمتعلّق بالحدث B. (ونرمز إليه بـ P { A/B}) ، العبارة :

$$P\{A/B\} = \frac{P\{A \cap B\}}{P\{B\}}$$

إنّ العبارة $\frac{P\{A \cap B\}}{P\{B\}}$ لها نفس طبيعة الاحتمال لأنّها تحقّق المبادىء الثلاثة المعروضة سابقاً :

(المبدأ 1) فهي بالفعل عدد إيجابي أو صفر ؛

(المبدأ 2) $P\{E/B\} = \frac{P\{E \cap B\}}{P\{B\}} = \frac{P\{B\}}{P\{B\}} = 1$

إذا أخذنا A و'A كحدثين متنافيين (الشكل 12) :

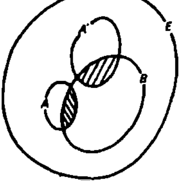
$$P\{A \cup A'/B\} = \frac{P\{(A \cup A') \cap B\}}{P\{B\}} = \frac{P\{(A \cap B) \cup (A' \cap B)\}}{P\{B\}}$$

(المبدأ 3) $$= \frac{P\{A \cap B\}}{P\{B\}} + \frac{P\{A' \cap B\}}{P\{B\}} = P\{A/B\} + P\{A'/B\}$$

إذا كان A وB حدثين باحتمالين لا يساويان صفراً ، نستنتج من التعريف المبدئي للاحتمال المشروط العلاقة المتقابلة التالية :

إنّ الإحتمال المشروط للحدث A والمتعلّق بالحدث B هو احتمال لتحقيق الحدث A عندما نعرف أنّ الحدث B قد تحقّق . ونقول :

P { A/B} : احتمال A إذا تحقّق B



الشكل 12

$$P\{A \cap B\} = P\{A\}.P\{B/A\} = P\{B\}.P\{A/B\}$$

تحمل هذه العلاقة اسم قاعدة الاحتمالات المركّبة ، وهي تسمح بحساب احتمال تحقيق حدثين في آن واحد .

مثل 1 : من وعاء يحتوي 10 كرات بيضاء ، 20 كرة حمراء و30 كرة سوداء نسحب كرتين دون أن تردّ الكرة المسحوبة إلى الوعاء . ما هو احتمال أن تكون الكرة الأولى المسحوبة حمراء والثانية بيضاء ؟ للحلّ طريقتان .

الطريقة الأولى : تعداد الحالات الممكنة والحالات المناسبة .

عدد الحالات الممكنة : هو عدد طرق اختيار كرتين مختلفتين إمّا باللون إما بترتيب السحب . إنّه عدد ترتيبات 60 كرة اثنين اثنين :

$$A_{60}^{2} = \frac{60\,!}{58\,!} = 59 \times 60$$

عدد الحالات المناسبة : هو عدد الأزواج (حمراء ، بيضاء) التي يمكننا تكوينها مع 20 كرة حمراء و10 كرات بيضاء ، أي 20×10=200 زوج . الاحتمال المطلوب هو إذن :

$$P\{\text{حمراء ، بيضاء}\} = \frac{200}{59 \times 60} = \frac{10}{177}.$$

الطريقة الثانية : تطبيق قاعدة الاحتمالات المركّبة

$$P\{\text{حمراء ، بيضاء}\} = P\{\text{حمراء}\}\, P\{\text{بيضاء / حمراء}\} = \frac{20}{60}.\frac{10}{59} = \frac{10}{177}$$

ففي الحقيقة ، الاحتمال المشروط للحصول على كرة بيضاء عند السحب الثاني ، مع العلم أننا حصلنا على كرة حمراء عند السحب الأوّل ، يساوي $\frac{10}{59}$: إذ بقي 59 كرة في الوعاء 10 منها بيضاء .

المثل 2 : إذا سحبنا ثلاث ورقات من ورق لعب (52 ورقة) ، دون ردّ الورقة

المسحوبة . ما هو احتمال الحصول على ثلاثة ملوك ؟

للحلّ أيضاً طريقتان .

الطريقة الأولى . تعداد الحالات الممكنة والحالات المناسبة .

عدد الحالات الممكنة : هو عدد طرق اختيار ثلاث ورقات ، دون أهمّية لطريقة الترتيب . إنّه عدد توافقيات ثلاث ورقات مُختارة بين 52 :

$$C_{52}^{3} = \frac{52\,!}{3\,!\,49\,!} = \frac{50 \times 51 \times 52}{2 \times 3}.$$

عدد الحالات المناسبة : هو عدد طرق اختيار ثلاثة ملوك ضمن مجموعة تتألّف من أربعة . إنّه عدد التوافقيات التي يمكننا إيجادها بواسطة الملوك الأربعة مأخوذة ثلاثة ثلاثة :

$$C_{4}^{3} = \frac{4\,!}{3\,!\,1\,!} = 4,$$

الاحتمال المطلوب هو إذن :

$$P\{\text{3 ملوك}\} = \frac{C_{4}^{3}}{C_{52}^{3}} = \frac{2 \times 3 \times 4}{50 \times 51 \times 52} = \frac{1}{5\,525}.$$

الطريقة الثانية : تطبيق قاعدة الاحتمالات المركّبة .

لنرمز بواسطة R_1 ، R_2 ، R_3 إلى مجموعات سحب ثلاث ورقات حيث يظهر ملك عند السحب الأول والثاني والثالث .

$$P\{R_1 \cap R_2 \cap R_3\} = P\{R_1\}.P\{R_2/R_1\}.P\{R_3/R_1 \cap R_2\}$$
$$= \frac{4}{52}.\frac{3}{51}.\frac{2}{50} = \frac{1}{5\,525}.$$

في الواقع، عند السحب الثاني، احتمال سحب ملك مع العلم أننا قد حصلنا على ملك عند السحب الأول يساوي $\frac{3}{51}$: إذ بقي 51 ورقة منها 3 ملوك . عند السحب الثالث ، لم يبق سوى 50 ورقة ، منها ملكان .

C ـ الاستقلالية بين حدثين

A وB هما حدثان باحتمالين مختلفين عن الصفر . نقول أنّ A مستقلّ عن B إذا كان :

$$P\{A/B\} = P\{A\}.$$

وهذا يعني أنّ احتمال تحقيق A لم يتأثر أبداً بكون B تحقّق أم لم يتحقّق . إذا عدنا إلى قاعدة الاحتمالات المركّبة :

$$P\{A \cap B\} = P\{A\}.P\{B/A\} = P\{B\}.P\{A/B\} = P\{B\}.P\{A\}.$$

نستنتج أنّ :

$$P\{B/A\} = P\{B\}.$$

الاستقلالية هي إذن خاصّة متبادلة : إذا كان A مستقلاً عن B ، B هو أيضاً مستقلّ عن A . بالتالي ، نقول أنّ A وB هما مستقلّان إذا امتازا بالعلاقة التالية :

$$P\{A \cap B\} = P\{A\}.P\{B\}.$$

هذه القاعدة هي قاعدة الاحتمالات المركّبة في حالة حدثين مستقلّين .

مثل 1 : إذا رمينا حجري زهر وأعطينا التفسير التالي لكل من الحدثين A وB :

A : الزهر الأوّل يعطي 1 ،

B : مجموع نقاط الزهرين هو مزدوج .

هل هذان الحدثان مستقلّان أم لا ؟

لنحسب $P\{A\}$ ، $P\{B\}$ و $P\{A \cap B\}$

$$P\{A\} = \frac{1}{6}$$

$$P\{B\} = P\{(P_1 \cap P_2) \cup (I_1 \cap I_2)\}$$

حيث P_1 وP_2 هما مجموعتا الحصول على عدد مزدوج على كل زهر ،

I_1 وI_2 هما مجموعتا الحصول على عدد مفرد على كلّ زهر .

فكي يكون مجموع النقاط على الزهرين مزدوجاً ، يجب أن تكون نقطتا الزهرين وفي آن واحد إمّا مزدوجتين ، إمّا مفردتين .

وإذا اعتمدنا قاعدة الاحتمالات الكلّية :

$$P\{B\} = P\{P_1 \cap P_2\} + P\{I_1 \cap I_2\}.$$

ولكن رمية كل زهر هي مستقلّة عن رمية الزهر الآخر :

$$P\{P_1 \cap P_2\} = P\{P_1\}.P\{P_2\}.$$

$$P\{I_1 \cap I_2\} = P\{I_1\}.P\{I_2\}.$$

وبما أنّنا نعرف قيمة كلّ احتمال :

$$P\{P_1\} = P\{P_2\} = P\{I_1\} = P\{I_2\} = \frac{1}{2}$$

$$P(B) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$$

من جهة أخرى :

$$P\{A \cap B\} = P\{1 \cap I_2\}.$$

ففي الواقع إذا حصلنا على 1 عند رمية الزهر الأوّل ، يجب أن نحصل على عدد مفرد عند رمية الزهر الثاني كي يصبح مجموع النقاط مزدوجاً . عندما نرمي زهرين، هناك 36 نتيجة ممكنة ومتعادلة الاحتمال من بينها 3 فقط تناسب الحدث$(1 \cap I_2)$ بالتالي :

$$P\{A \cap B\} = \frac{1}{12}.$$

إذن

$$P\{A \cap B\} = P\{A\}.P\{B\}$$

$$\frac{1}{12} = \frac{1}{6} \cdot \frac{1}{2}.$$

الحدثان A و B هما إذن مستقلّان .

مثل 2 : رمينا قطعة نقود n مرة وأعطينا التفسير التالي للحدثين A وB :

A : نحصل على الجهة face مرّة واحدة على الأكثر ؛

B : نحصل على كل من الجهتين pile وface على الأقلّ مرّة واحدة . هل الحدثان A وB مستقلّان ؟

النتيجة تكون حسب عدد الرميات n .

إذا كان n = 2 ، فإن كلّ الإمكانيات المحتملة والمتعادلة الاحتمال هي :

FF, FP, PF, PP

الإمكانيات التي تنتج الحدث A هي : PF, FP وPP

الحدث B : FP وPF

والحدث $A \cap B$: FP وPF . بالتالي :

$$P\{A\} = \frac{3}{4}, \qquad P\{B\} = \frac{1}{2} \quad \text{et} \quad P\{A \cap B\} = \frac{1}{2}$$

إذن الحدثان A وB ليسا مستقلّين .

إذا كان n = 3 ، فإنّ كلّ الإمكانيات المحتملة والمتعادلة الاحتمال هي :

FFF, FFP, FPF, PFF

FPP, PFP, PPF, PPP

الإمكانيات التي تنتج الحدث A هي : PPP وPPF, PFP, FPP

الحدث B : PPF وPFP, FPP, PFF, FPF, FFP

والحدث $A \cap B$: PPF وPFP, FPP . بالتالي :

$$P\{A\} = \frac{1}{2}, \qquad P\{B\} = \frac{3}{4} \quad \text{و} \quad P\{A \cap B\} = \frac{3}{8}.$$

إذن :

$$P\{A \cap B\} = P\{A\}.P\{B\}$$
$$\frac{3}{8} = \frac{1}{2} \cdot \frac{3}{4}$$

الحدثان A وB هما إذن مستقلّان .

إنّ قواعد الحساب التي درسناها لتوّنا في الحالة حيث مجموعة الأحداث متناهية تبقى صالحة إذا كانت هذه المجموعة غير متناهية ويمكن تعداد عناصرها أو غير متناهية ولا يمكن تعداد عناصرها . سوف نلتقي خلال دراستنا للمتغيّرات العشوائية (الصدفية) ولقوانين الاحتمال بأمثلة عن مجموعات من هذا النوع .

إلاّ أنّه عندما تكون مجموعة الأحداث غير متناهية ، يجب وضع مبدأ إضافي يبسط المبدأ 3 إلى عدد غير متناه من الأحداث :

3َ : إنّ احتمال اتحاد سلسلة غير متناهية وممكنة التعداد من الأحداث A_i حيث يتنافى كلّ حدث مع الآخر يساوي المجموع غير المتناهي لاحتمالات هذه الأحداث :

$$P\left\{ \bigcup_{i=1}^{\infty} A_i \right\} = \sum_{i=1}^{\infty} P\{A_i\}$$

من ناحية أخرى :

عندما تكون مجموعة الأحداث E متناهية ، أو غير متناهية ولكن يمكن تعداد عناصرها ، فإنّ الاحتمال يتحدّد على مجموعة أجزاء E أي $\wp(E)$:

نسب احتمالاً إلى كلّ جزء من E .

بالمقابل ، عندما تكون مجموعة الأحداث غير متناهية ولا يمكن تعداد عناصرها ،

مثلاً ، مجموعة نقاط خط مستقيم أو نقاط مسطّح ما)، لا يمكن تحديد احتمال على مجمل المجموع $\mathcal{P}(E)$ يحقّق المبادىء السابقة . هنا نضطر أن نحصر تحديد الاحتمال على عائلة F من أجزاء المجموعة E . ويجب أن تكون لهذه العائلة نفس البنية التي كانت لمجموعة أجزاء E أي $\mathcal{P}(E)$ في الحالات السابقة ، أي أنّها يجب أن تفي بالشروط التالية :

أ - إذا كان الحدث A عنصراً من F ، فإنّ متمم A بالنسبة للمجموعة E ينتمي أيضاً إلى F ،

ب - إذا كان الحدثان A وB عنصرين من F ، فإنّ $A \cup B$ و $A \cap B$ ينتميان أيضاً إلى F ؛

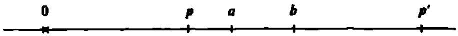
ج - كل اتحاد يمكن تعداده بين عناصر A_i من F هو أيضاً عنصر من F :

$$\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}.$$

إنّ الشرطين الأوّلين اللذين يحقّقها P (E) عندما تكون مجموعة الأحداث متناهية ، يحدّدان ما يسمّى بجبر بول (algèbre de Boole) . والشرط الثالث كان ضرورياً لأنّ مجموعة الأحداث غير متناهية : وهو يبسط الشرط ب إلى عدد غير متناه من الأحداث ، وتحقّقه المجموعة P (E) عندما تكون مجموعة الأحداث غير متناهية ولكن يمكن تعداد عناصرها . كل هذه الشروط تحدّد ما يُسمّى σ - جبر (سيغما جبر ، σ -algèbre) أو عائلة بوريل (famille de Borel) .

من أجل تحديد احتمال عندما تكون مجموعة الأحداث غير متناهية ولا يمكن تعداد عناصرها ، نضطر إذن لاستبدال مجموعة أجزاء E أي $\mathcal{P}(E)$ بعائلة من أجزاء E تشكّل σ - جبر .

مثلاً : لنأخذ عشوائياً نقطة على قطعة المستقيم ′pp :



إنّ مجموعة الأحداث المنسوبة إلى هذه التجربة هي مجموعة نقاط القطعة ′pp وهي مجموعة غير متناهية ولا يمكن تعداد عناصرها . كلّ نقطة من هذه القطعة لها نفس الاحتمال لأن تؤخذ كجاراتها ،وبما أن هناك عدداً غير متناه من النقاط ، هذا الاحتمال يساوي صفراً . نعرف إذن ، بشكل خاص ، أنّ الاحتمال المنسوب إلى كل نقطة من المسافة (a,b) يساوي صفراً . ولكن ليس من الممكن ، انطلاقاً من المبادىء 1 ، 2 و′3

السابقة، استنتاج احتمال المسافة (a, b) (أي احتمال أنْ تكون النقطة المأخوذة تنتمي إلى هذه المسافة) .

بالمقابل ، من الطبيعي أن نعطي المسافة (a, b) احتمالاً يساوي نسبة طول هذه المسافة على طول القطعة pp' الإجمالي : $P(a, b) = \frac{b - a}{p' - p}$

ويحقّق هذا التحديد المبادىء السابقة .

نرى إذن أنّه من الضروري المرور بواسطة المسافات (a, b) لتحديد احتمال بالنسبة لقطعة من المستقيم . هذه المسافات تولّد ،بواسطة العمليات أ ،ب وج - جبر F . وهكذا بالإمكان تحديد احتمال كل عنصر من F . لنشر أنه في هذه الحالة الخاصّة ، كلّ جزء من القطعة pp' يتكوّن من نقطة واحدة احتماله يساوي صفراً ، وكذلك كلّ جزء يتكوّن من عدد متناه من النقاط أو أيضاً من عدد غير متناه من النقاط ولكن يمكن تعداده ينتمي إلى F واحتماله يساوي صفراً .

القسم IV

مفهوم المتغيّرة العشوائية وقانون الاحتمال

1 . المتغيّرات العشوائية وقوانين الاحتمال ذات البعد الواحد : A. تعريفات ؛ B. المتغيّرات المنفصلة ؛ C. المتغيّرات المتواصلة .ـ 2. المتغيّرات العشوائية وقوانين الاحتمال ذات البعدين : A. تعريف ؛ B . المتغيّرات المنفصلة ؛ C. المتغيّرات المتواصلة .

1 . المتغيرات العشوائية وقوانين الاحتمال ذات البعد الواحد

A . تعريفات

نحدّد متغيّرة عشوائية X عندما ننسب عدداً معيناً إلى كلّ حدث نموذجي من مجموعة الأحداث E .

وإذا نسبنا لكلّ قيمة ممكنة من قيم المتغيّرة العشوائية ، احتمال الحدث المطابق لها نحصل على قانون الاحتمال (أو توزيع الاحتمال) للمتغيّرة العشوائية X .

مثل 1 . نرمي مرّتين على التوالي قطعة من النقود ونحدّد المتغيّرة العشوائية X بعدد المرات التي نحصل فيها على الوجه face خلال هاتين الرميتين . عندئذٍ نحصل على قانون الاحتمال التالي (القراءة من اليسار إلى اليمين) .

الحدث النموذجي	المتغيّرة العشوائية X	الاحتمال $P\{X\}$
$P_1\ P_2$	0	1/4
$P_1\ F_2$ $F_1\ P_2$	1	1/2
$F_1\ F_2$	2	1/4
المجموع		1

حيث P_1 ترمز إلى الحصول على الوجه pile عند الرمية الأولى ، P_2 الحصول على الوجه pile عند الرمية الثانية ، F_1 الحصول على الوجه face عند الرمية الأولى وF_2 الحصول على الوجه face عند الرمية الثانية .

بوسع المتغيّرة العشوائية X أن تأخذ القيم 0 ,1 و2 ، وهذه القيم تكوّن ما يُسمّى مجموعة تحديد المتغيّرة .

مثل 2 . من وعاء يحتوي على كرات بيضاء بنسبة p وكرات حمراء بنسبة q ، (q=1−p) نسحب بالصدفة كرة واحدة . نحدّد المتغيّرة العشوائية X بالطريقة التالية : X=1 إذا كانت الكرة المسحوبة بيضاء وX=0 إذا كانت حمراء ، فنحصل على قانون الاحتمال التالي (القراءة من اليسار إلى اليمين) :

الحدث النموذجي	المتغيّرة العشوائية X	الاحتمال P { X }
B (بيضاء)	1	p
R (حمراء)	0	$q = 1 - p$
المجموع		1

تُسمّى المتغيّرة العشوائية المحدّدة بهذه الطريقة متغيّرة برنولي (Bernouilli) ، ومجموعة تحديدها هي {0, 1} .

سوف نستعملها في الفصل II لدراسة القانون ذي الحدّين (binomial) .

إن حاصل جمع الاحتمالات التي تؤلّف قانون الاحتمال يساوي دائماً واحداً ، فهو في الواقع يساوي مجموع احتمالات كلّ الأحداث النموذجية .

هذه التعريفات يجب أن تتعدّل بعض الشيء عندما تكون مجموعة الأحداث E غير متناهية ولا يمكن تعداد عناصرها .

في الحقيقة ، عندما تكون E غير متناهية ولا يمكن تعداد عناصرها ، لا يعود ممكناً تحديد احتمال على أي جزء من E ، إذ يجب أن نحصر الأمر بعائلة F من أجزاء E تشكّل σ - جبر .

لنأخذ التخصيص الرقمي الذي ينسب إلى كل عنصر من E عدداً حقيقياً . إنّ المجموعة التي تكوّن صورة E هي جزء ما من مجموعة الأعداد الحقيقية R ، لا يمكننا إذن تحديد احتمال على هذه المجموعة غير المتناهية والتي لا يمكن تعداد عناصرها إلاّ باعتمادنا عناصر من σ - جبر من R ، مثلاً المسافات المفتوحة $]-\infty, x[$. كذلك يجب أن نعرف كيف نخصِّص احتمالاً لِـ A_x وهي الصورة المعكوسة للمسافة $]-\infty, x[$. لهذا ، من الضروري أن تنتمي A_x إلى الـ σ - جبر F الذي يتشكّل من أجزاء من E . إذن التحديد العام للمتغيِّرة العشوائية هو التالي :

P هو احتمال محدّد على عائلة الأجزاء F التي تؤلّف σ - جبر، التخصيص الرقمي X الذي ينسب إلى كلّ عنصر من E عدداً حقيقياً x ، هو متغيِّرة عشوائية إذا كانت الصورة المعكوسة A_x ، مهما كان x ، للمسافة المفتوحة $]-\infty, x[$ تنتمي إلى F . وتُسمّى المجموعة التي تكوّن صورة E بواسطة التخصيص X مجموعة تحديد المتغيِّرة العشوائية X .

التخصيص الذي ينسب إلى كلّ مسافة $]-\infty, x[$ احتمال الجزء المطابق A_x من مجموعة الأحداث هو وظيفة التوزيع $F(x)$ للمتغيِّرة العشوائية X :

$$F(x) = P(X < x) = P(A_x).$$

نسمّي وظيفة توزيع المتغيِّرة العشوائية X ، الوظيفة العددية (الرقمية) الإيجابية F التالية :

$$F(x) = P(X < x).$$

وهي احتمال أن تكون المتغيِّرة العشوائية X أصغر من قيمة معيّنة x .

دلالات . بشكل عام ندلّ بواسطة X (أو Y ، أو Z ، . . .) على متغيِّرة عشوائية وبواسطة x (أو y ، أو z ، . . .) على قيمة معيّنة لهذه المتغيِّرة .

ونميِّز بين المتغيِّرات العشوائية المنفصلة (حيث مجموعة التحديد متناهية أو غير متناهية ولكن يمكن تعداد عناصرها) والمتغيِّرات العشوائية المتواصلة (حيث مجموعة التحديد غير متناهية ولا يمكن تعداد عناصرها) . هنا نجد تصنيفاً مشابهاً لما صادفناه بالنسبة للمتغيِّرات الإحصائية ، كما نشير من جهة أخرى إلى التشابه الحاصل ، بشكل

عام ، بين المتغيّرات العشوائية والمتغيّرات الإحصائية ، حيث يحل مفهوم الاحتمال بالنسبة للمتغيّرات العشوائية محل مفهوم التردد (التكرار) بالنسبة للمتغيّرات الإحصائية : الاحتمال هو التردد المثالي الذي يطابق عدداً غير متناه من الحالات الملحوظة . وسيسمح لنا قانون الأعداد الكبيرة الذي سندرسه في الفصل V بإقامة جسر بين هذين المفهومين .

B . المتغيّرات المنفصلة

نقول أنّ المتغيّرة X هي منفصلة إذا كان عدد مختلف قيمها الممكنة متناهياً أو غير متناه ولكن يمكن تعدادها .

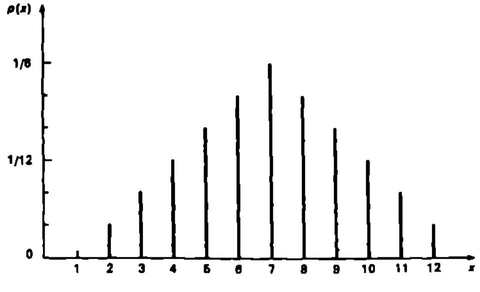
قانون الاحتمال

ينسب قانون الاحتمال إلى كلّ قيمة ممكنة للمتغيّرة المنفصلة X احتمال الحدث المطابق . التمثيل البياني له هو مخطّط العيدان .

مثل 1 . نرمي حجري زهر ونحدّد المتغيّرة العشوائية X وهي عبارة عن حاصل جمع نقاط الحجرين .

مجموعة القيم الممكنة ، أو مجموعة تحديد المتغيّرة العشوائية X ، هي المجموعة $\{2, 3, \ldots, 12\}$ ، إنّها مجموعة متناهية .

نحصل على قانون الاحتمال التالي ، وتمثيله البياني في الشكل 13 (القراءة من اليسار إلى اليمين) :



الشكل 13 . مخطّط العيدان (المثل 1) .

الحدث النموذجي	المتغيّـرة العشـوائية X	الاحتمال P { X }
1,1	2	1/36
1,2 2,1	3	1/18
1,3 2,2 3,1	4	1/12
1,4 2,3 3,2 4,1	5	1/9
1,5 2,4 3,3 4,2 5,1	6	5/36
1,6 2,5 3,4 4,3 5,2 6,1	7	1/6
2,6 3,5 4,4 5,3 6,2	8	5/36
3,6 4,5 5,4 6,3	9	1/9
4,6 5,5 6,4	10	1/12
5,6 6,5	11	1/18
6,6	12	1/36
Total		1

مثل 2 . نرمي قطعة من النقود ونحدّد المتغيّرة العشوائية X وهي عبارة عن عدد الرميات المتالية الضرورية قبل الحصول على الجهة pile للمرّة الأولى : مجموعة القيم الممكنة {x} هي مجموعة الأعداد الصحيحة الإيجابية :

$$\{ x \} = \{ 1, 2, 3, ... \}$$

وهي مجموعة غير متناهية ولكن يمكن تعداد عناصرها .

كي تكونx رمية ضرورية ، يجب الحصول على الجهة face عند الرميات (x−1) الأولى والجهة pile عند الرمية رقم x ، إذن :

$$P \{ X = x \} = \left(\frac{1}{2}\right)^{x-1} \times \frac{1}{2} = \frac{1}{2^x}.$$

ونحصل على قانون الاحتمال التالي (القراءة من اليسار إلى اليمين) :

المتغيّرة العشوائية X	الاحتمال $P\{X\}$
1	1/2
2	1/4
3	1/8
⋮	⋮
x	$1/2^x$
⋮	⋮
المجموع	1

بإمكاننا التثبّت من أن مجموع الاحتمالات يساوي واحداً[1] .

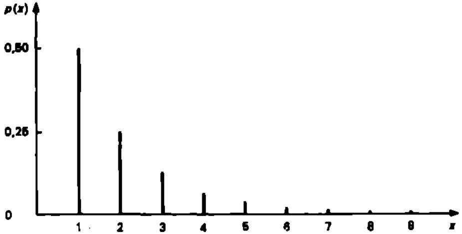
الشكل 14 هو التمثيل البياني لهذا القانون :

(1) إنّه حاصل جمع متوالية هندسية لا متناهية ، المجموع :

$$S = a + aq^1 + aq^2 + \cdots,$$

حيث العنصر الأوّل هو a = 1/2 والأساس هو q = 1/2 (أصغر من 1) :

$$S = \frac{a}{1-q} = 1 .$$



الشكل 14 . مخطط الميدان (المثل 2)

وظيفة التوزيع

وظيفة توزيع المتغيّرة المنفصلة X ، المحدّدة بواسطة :

$$F(x) = P\{X < x\}$$

هي وظيفة إيجابية غير تنازلية .

وتساوي هذه الوظيفة صفراً عند ∞ - :

حدّ $0 = F(x)$ عندما $x \to -\infty$

وواحداً عند ∞ + :

حدّ $1 = F(x)$ عندما $x \to +\infty$

عندما تكون مجموعة القيم الممكنة ، أو مجموعة تحديد المتغيّرة العشوائية X متناهية :

$$\{x_1, x_2, \ldots, x_i, \ldots, x_n\},$$

فإن F (x) تساوي صفراً على الفسحة $]-\infty, x_1[$ وتساوي واحداً على الفسحة $]x_n, +\infty[$.

وتحتفظ وظيفة التوزيع بنفس القيمة $F(x_i)$ على كل فسحة $]x_i, x_{i+1}[$ ، وعند النقطة ذاتَ الإحداثية السينية x_i تقوم بقفزة تساوي الاحتمال المنسوب إلى القيمة x_i .

يمكننا بسهولة حساب وظيفة التوزيع انطلاقاً من الاحتمالات المنسوبة إلى القيم الممكنة للمتغيّرة المنفصلة :

$$F(x) = \sum_{x_i < x} P\{ X = x_i \} .$$

وبالعكس تسمح لنا وظيفة التوزيع بإيجاد توزيع الاحتمالات :

$$P\{ X = x_i \} = F(x_{i+1}) - F(x_i) .$$

إذن لا يهمّ أن يكون لدينا وظيفة التوزيع أم قانون الاحتمال .

التمثيل البياني لوظيفة التوزيع هو المنحنى التراكمي . في حالة المتغيّرة المنفصلة ، نسميه أيضاً المنحنى ـ الدرج وذلك لشكله ، فهو عبارة عن درجات (أو قفزات) عند النقاط ذات الإحداثيات السينيات (abscisses) x_i التي تطابق القيم الممكنة للمتغيّرة .

وظيفة التوزيع هي بالنسبة للمتغيرات العشوائية ، ما يعادل وظيفة التردد (fréquence) التراكمية بالنسبة للمتغيّرات الإحصائية .

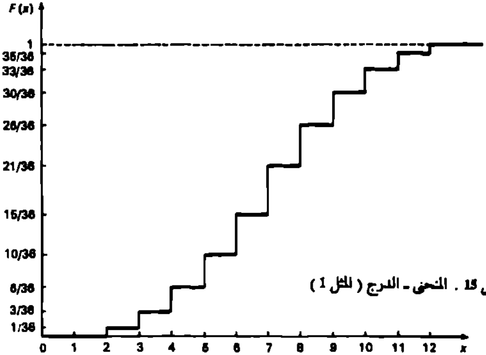
لنعد إلى المثلين السابقين .

مثل 1 . وظيفة توزيع المتغيّرة العشوائية المحدّدة كمجموع النقاط الحاصلة على الزهرين هي التالية (القراءة من اليسار إلى اليمين) :

المتغيّرة العشوائية X	الاحتمال $P\{X\}$	وظيفة التوزيع $F(X)$
		0
2	1/36	
		1/36
3	1/18	
		3/36
4	1/12	
		6/36
5	1/9	
		10/36
6	5/36	
		15/36
7	1/6	
		21/36
8	5/36	
		26/36
9	1/9	
		30/36
10	1/12	

		33/36
11	1/18	
		35/36
12	1/36	
		1
Total	1	

التمثيل البياني لوظيفة التوزيع هذه هو المنحنى ـ الدرج المقدّم في الشكل 15 .



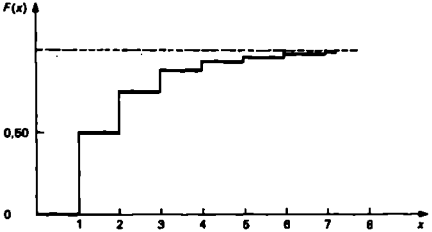
الشكل 15 . المنحنى ـ الدرج (المثل 1)

مثل 2 . وظيفة توزيع المتغيّرة العشوائية المحدّدة كعدد رميات قطعة النقود

المتغيّرة العشوائية X	الاحتمال $P(X)$	وظيفة التوزيع $F(X)$
		0
1	1/2	
		1/2
2	1/4	
		3/4
3	1/8	
⋮	⋮	7/8
		⋮
x	$1/2^x$	
		$(2^x - 1)/2^x$
⋮	⋮	⋮
المجموع	1	1

الضرورية قبل الحصول على الجهة pile هي واردة في الجدول (القراءة من اليسار إلى اليمين.) .

وتمثيلها البياني هو المنحنى ـ الدرج المقدّم في الشكل 16 .



الشكل 16 . المنحنى ـ الدرج (المثل 2)

C . المتغيّرات المتواصلة

نقول أنّ المتغيرة العشوائية X هي متواصلة إذا كانت مجموعة تحديدها عبارة عن فسحة .

وظيفة التوزيع

يُحدّد توزيع احتمال متغيّرة عشوائية متواصلة بواسطة وظيفة التوزيع :

$$F(x) = P\{X < x\}.$$

F(x) هي وظيفة إيجابية تصاعدية (متزايدة) ، كما أنّـه :

$$\lim_{x \to -\infty} F(x) = 0, \qquad \lim_{x \to +\infty} F(x) = 1.$$

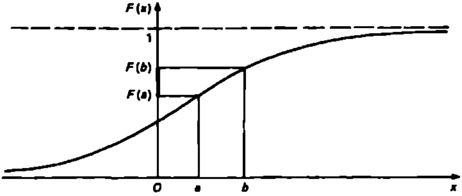
أي أنّ حدّ F(x) يساوي صفراً عند ∞ − وواحداً عند ∞ + .

إذا كانت الوظيفة F(x) متواصلة ولها مشتقّة f(x) ، نقول أنّ المتغيّرة X هي متواصلة مطلقاً .

المنحنى التراكمي أو منحنى التوزيع هو التمثيل البياني لوظيفة التوزيع F(x) (الشكل 17) .

الاحتمال المنسوب إلى فسحة

يساوي احتمال أن تنتمي X إلى الفسحة (a, b) الفارق بين القيمتين اللتين



الشكل 17 . منحنى التوزيع

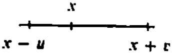
تأخذها وظيفة التوزيع عند طرفي الفسحة :

$$P\{a \leqslant X < b\} = P\{X < b\} - P\{X < a\} = F(b) - F(a).$$

الاحتمال المنسوب إلى نقطة

عندما تكون المتغيِّرة X متواصلة مطلقاً ، يكون الاحتمال المنسوب إلى النقطة x صفراً .

في الواقع ، لنأخذ العددين الإيجابيين u وv ، النقطة x تنتمي إلى الفسحة (x + v) ،(x−u .



يمكننا الكتابة :

$$0 \leqslant P\{X = x\} \leqslant P\{x - u \leqslant X < x + v\}$$
$$0 \leqslant P\{X = x\} \leqslant F(x + v) - F(x - u)$$
$$0 \leqslant P\{X = x\} \leqslant [F(x + v) - F(x)] + [F(x) - F(x - u)].$$

وبما أنّ F(x) هي وظيفة متواصلة :

$$[F(x + v) - F(x)] \to 0 \quad \text{إذا} \quad v \to 0$$
$$[F(x) - F(x - u)] \to 0 \quad \text{إذا} \quad u \to 0$$

إذن

$$\underline{P}\{X = x\} = 0.$$

كثافة الاحتمال عند نقطة معيّنة

الاحتمال المنسوب إلى الفسحة (a, b) هو :

$$P\{a \leqslant X < b\} = P\{X < b\} - P\{X < a\} = F(b) - F(a).$$

كثافة الاحتمال المتوسّطة على الفسحة (a, b) هي نسبة هذا الاحتمال على طول الفسحة :

$$f(a, b) = \frac{F(b) - F(a)}{b - a}.$$

∴ بالتالي ، الكثافة المتوسّطة للاحتمال على فسحة صغيرة (x, x + Δx) هي :

$$f(x, x + \Delta x) = \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

نسمّي كثافة الاحتمال f(x) عند نقطة x ، القيمة الحدّ للكثافة المتوسّطة على المسافة (x, x + Δx) عندما يميل طول هذه الفسحة Δx إلى الصفر :

$$f(x) = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x} = F'(x).$$

. (حدّ = lim)

إذن كثافة الاحتمال هي مشتقّة وظيفة التوزيع . وتمثيلها البياني هو منحنى كثافة الاحتمال (الشكل 18) .

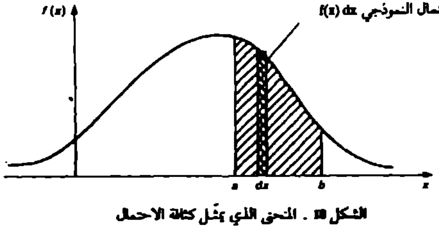
الاحتمال النموذجي لأن تأخذ المتغيّرة العشوائية X قيمة داخل فسحة لا متناهية الصغر بطول dx لحل بالتقـ ... ضرب كثافة الاحتمال بطول الفسحة :

$$P\{x \leqslant X < x +$$

الاحتمال المنسوب إلى الفسحة (a, b) يبدو إذن كأنه مجموع هذه الاحتمالات النموذجية مأخوذاً بين a وb :

$$P\{a \leqslant X < b\} = \int_a^b f(x)\,dx = F(b) - F(a).$$

نمثّل هذا الاحتمال في الشكل 18 بواسطة المساحة المخطّطة :



الشكل 18 . المنحني الذي يمثّل كثافة الاحتمال

المساحة المحصورة بين منحنى كثافة الاحتمال ومحور الإحداثيات السينيات (abscissas) تساوي واحداً لأن :

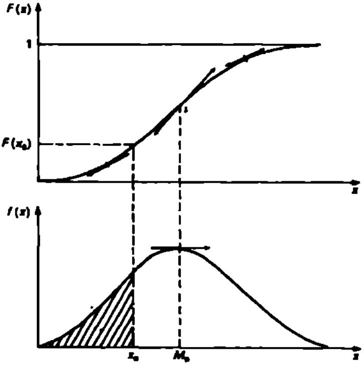
$$\int_{-\infty}^{+\infty} f(x)\, dx = F(+\infty) - F(-\infty) = 1$$

الشكل 19 يعرض العلاقات الموجودة بين وظيفة التوزيع وكثافة الاحتمال . نعبر من كثافة الاحتمال إلى وظيفة التوزيع كما نعبر ، بالنسبة للمتغيّرات الاحصائية من المدرج التكراري إلى منحني التردّد التراكمي . قيمة وظيفة التوزيع F(x0) هي مجموع كلّ الاحتمالات النموذجية المطابقة للقيم x الأصغر من x0 . إذن F(x0) تساوي المساحة المخطّطة المحصورة بين منحنى كثافة الاحتمال ومحور الإحداثيات السينيات، أي ما نرمز إليه بواسطة :

$$F(x_0) = \int_{-\infty}^{x_0} f(x)\, dx$$

إذا كانت كثافة الاحتمال $f(x)$ وظيفة متواصلة ، وذات مشتقّة أولى $f'(x)$ ومشتقّة ثانية $f''(x)$ ،فإن منوال (mode) منحنى الكثافة Mo يطابق :

$$f'(M_0) = 0 , \qquad f''(M_0) < 0 .$$



الشكل 19 . العلاقة بين وظيفة التوزيع وكثافة الاحتمال

أي أنّه ، بالنسبة لوظيفة التوزيع :

$$F''(M_0) = 0 , \qquad F'''(M_0) < 0 .$$

تشير العلاقتان الأخيرتان ، إلى وجود نقطة انعطاف . إذن يطابق منوال منحنى الكثافة نقطة الانعطاف في المنحنى التراكمي (منحنى وظيفة التوزيع) . بالنسبة للقيم x الأصغر من Mo ، يتصاعد المنحنى التراكمي بسرعة أكثر فأكثر ، وهذا ما يُترجَم بمماس يوجد تحت المنحنى . بعد Mo ، يبقى المنحنى آخذاً في التصاعد ولكن بسرعة تصغر تدريجياً : عندها يكون المماس موجوداً فوق المنحنى : نقطة الانعطاف ، ذات الإحداثي السيني Mo ، هي النقطة حيث المماس يخترق المنحنى .

2 . المتغيّرات العشوائية وقوانين الاحتمال ذات البعدين

A . تعريف

لنفترض أن X وY هما متغيّرتان عشوائيتان محدّدتان على مجموعة الأحداث E .
إذا نسبنا على كلّ قيمة ممكنة للزوج (X, Y) احتمال الحدث المطابق فإننا نحصل على القانون الموصول للمتغيّرتين X وY، أو قانون المتغيرة العشوائية ذات البعدين (X,Y) .

v.a. X \ v.a. Y	2	3	4	5	6	7	8	9	10	11	12	قانون X الهامشي
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0	0	$\frac{1}{6}$
2	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	0	$\frac{1}{6}$
3	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	0	$\frac{1}{6}$
4	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	0	$\frac{1}{6}$
5	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	0	$\frac{1}{6}$
6	0	0	0	0	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{6}$
قانون Y الهامشي	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$	1

مثلاً . نرمي حجري زهر ونحدّد المتغيّرة العشوائية X كعدد النقاط الحاصلة على الزهر الأول والمتغيّرة العشوائية Y كمجموع نقاط الحجرين .

نحصل عندها على قانون الاحتمال ذي البعدين في الجدول أعلاه (القراءة من اليسار إلى اليمين) .

كما في حالة المتغيّرات العشوائية ذات البعد الواحد ، هذه التعريفات تتعدّل بعض الشيء عندما تكون مجموعة الأحداث غير متناهية ولا يمكن تعداد عناصرها .

B . المتغيّرات المنفصلة

لنرمز بواسطة p_{ij} إلى احتمال أن تأخذ X وY قيمتين معيّنتين x_i وy_j :

$$p_{ij} = P\{ X = x_i, Y = y_j \} .$$

إذن

$$\sum_i \sum_j p_{ij} = 1 .$$

أي أنّ مجموع الاحتمالات المنسوبة إلى القيم الممكنة للزوج (X, Y) يساوي واحداً. لنرمز بواسطة $p_{i.}$ إلى حاصل جمع الاحتمالات p_{ij} حسب الدليل j (أنظر كتاب « الاحصاء الوصفي » ، الفصل III ، القسم I) :

$$p_{i.} = \sum_j p_{ij} = P\{ X = x_i \} .$$

الاحتمالات $p_{i.}$ تكوّن قانون الاحتمال الهامشي للمتغيّرة X ،

كذلك عندما نجمع الاحتمالات p_{ij} حسب الدليل i :

$$p_{.j} = \sum_i p_{ij} = P\{ Y = y_j \} .$$

الاحتمالات $p_{.j}$ تكوّن قانون الاحتمال الهامشي للمتغيّرة Y .

v.a. Y / v.a. X	y_1	....	y_j	....	قانون X الهامشي
x_1	p_{11}	....	p_{1j}	....	$p_{1.}$
⋮	⋮		⋮		⋮
x_i	p_{i1}	....	p_{ij}	....	$p_{i.}$
⋮	⋮		⋮		⋮
قانون Y	$p_{.1}$	....	$p_{.j}$	....	1

في المثل السابق ، وجدنا قانون الاحتمال الهامشي اللذي يعطي توزيع مجموع نقاط الزهرين ، وهو قانون سبق أن حسبناه . الطريقة الحاضرة تعطينا وسيلة سهلة لإيجاده .

لنفترض أنّ احتمال أن تأخذ X القيمة x_i يختلف عن الصفر : $p_{i.} \neq 0$، الاحتمال المشروط لـ $Y=y_j$ مع العلم أ ن $X=x_i$ يساوي :

$$p_{j/i} = P\{ Y = y_j/X = x_i \} = \frac{p_{ij}}{p_{i.}}.$$

الاحتمالات $p_{j/i}$ المنسوبة إلى مختلف قيم Y الممكنة تكوّن القانون المشروط للمتغيّرة Y متعلّقاً بـ $X=x_i$.

كذلك ، إذا كانت $p_{.j} \neq 0$ ، الاحتمال المشروط لـ $X=x_i$ مع العلم أ ن $Y=y_j$ يساوي :

$$p_{i/j} = P\{ X = x_i/Y = y_j \} = \frac{p_{ij}}{p_{.j}}.$$

الاحتمالات $p_{i/j}$ المنسوبة إلى مختلف قيم X الممكنة تكوّن القانون المشروط للمتغيّرة X متعلّقاً بـ $Y = y_j$.

لقد سبق أن التقينا ، في ما يخصّ احتمال تحقيق حدثين في آن واحد ، بفكرة الاحتمال المشروط . بشكل خاص العلاقة التالية الموجودة ، وذلك بسبب التعريفات السابقة ، بين الاحتمالات الهامشية والمشروطة :

$$p_{ij} = p_{i.} \times p_{j/i} = p_{.j} \times p_{i/j}$$

تطابق قاعدة الاحتمالات المركّبة (انظر القسم III ، الفقرة 2.B) .

الاستقلالية

لنبسط تعريف الاستقلالية بين حدثين إلى المتغيرتين العشوائيتين X وY (أنظر القسم III ، الفقرة 2.C) :

نقول أنّ المتغيّرتين X وY هما مستقلّتان إذا حقّقتا العلاقة :

$$p_{ij} = p_{i.} \times p_{.j}$$

مهما كانت قيمة الزوج (x_i, y_j) ، أي أنّه مهما كان x_i وy_j ، الحدثان $(X=x_i)$ و$(Y=y_j)$ هما مستقلّان .

في هذه الحالة تتساوى الاحتمالات المشروطة مع الاحتمال الهامشي المناسب :

$$p_{ij} = \frac{p_{ij}}{p_{i.}} = p_{.j} , \qquad p_{ij} = \frac{p_{ij}}{p_{.j}} = p_{i.} .$$

وهذا يعني أنّ معرفة القيمة التي تأخذها X لا تحمل أي معلومات عن قيمة Y ، والعكس بالعكس .

إنّ قانون احتمال المتغيّرة العشوائية ذات البعدين (X, Y) يسمح لنا دون شك بحساب قانوني الأحتمال الهامشيين للمتغيّرتين X وY . ولكن بالمقابل ، معرفة هذين القانونين لا تسمح لنا بتحديد القانون الموصول ، إلاّ إذا كانت المتغيّرتان X وY مستقلّتين .

سوف نلاحظ وجه الشبه الحاصل بين مفهومي قوانين الاحتمال الهامشية والمشروطة لمتغيّرة عشوائية وقوانين التوزيعات الهامشية والمشروطة لمتغيّرة إحصائية (أنظر كتاب « الإحصاء الوصفي » ، الفصل III ، القسم I) . وقد ازدادت أدوات التحليل التي بحوزتنا غنى بإدخال فكرة الاستقلالية .

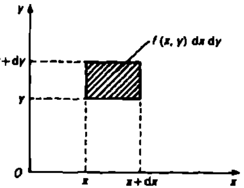
C . المتغيّرات المتواصلة

يوجد بالنسبة للمتغيّرات المتواصلة ذات البعدين تعريفات وخصائص شبيهة بالتي درسناها لتوّنا في حالة المتغيّرات المنفصلة .

إنّ الاحتمال النموذجي كي تأخذ المتغيّرة العشوائية (X, Y) قيمة داخل المستطيل اللامتناهي الصغر وذي المساحة dxdy الذي يحيط بالنقطة (x, y) (الشكل 20) هو :

$$P\{ x \leqslant X < x + \mathrm{d}x, y \leqslant Y < y + \mathrm{d}y \} = f(x, y)\, \mathrm{d}x\, \mathrm{d}y ,$$

حيث f(x, y) تمثّل كثافة احتمال المتغيّرة ذات البعدين .



الشكل 20

ويساوي مجموع كل الاحتمالات النموذجية واحداً :

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y)\, \mathrm{d}x\, \mathrm{d}y = 1 .$$

أمّا الكثافتان الهامشيتان للمتغيّرتين فهما على التوالي :

$$f(x) = \int_{-\infty}^{+\infty} f(x, y)\, dy , \qquad f(y) = \int_{-\infty}^{+\infty} f(x, y)\, dx .$$

كثافة الاحتمال المشروطة للمتغيّرة Y متعلّقة بـ X = x هي :

$$g(y/x) = \frac{f(x, y)}{f(x)} = \frac{f(x, y)}{\int_{-\infty}^{+\infty} f(x, y)\, dy} .$$

كذلك ، كثافة الاحتمال المشروطة للمتغيّرة X متعلّقة بـ X = y هي :

$$h(x/y) = \frac{f(x, y)}{f(y)} = \frac{f(x, y)}{\int_{-\infty}^{+\infty} f(x, y)\, dx}$$

أخيراً ، نقول أنّ المتغيّرتين X وY مستقلتان إذا حقّقتا العلاقة التالية :

$$f(x, y) = f(x).f(y) .$$

مهما كانت قيمة الزوج (x, y) .

القسم V

مقاييس المتغيّرة العشوائية

1 . الأمل الرياضي : A . تعريف ؛ B . خصائص .ـ 2 . التباين : A . تعريف ؛ B . خصائص ـ 3 . تغاير متغيّرتين عشوائيتين . 4 . العزم .

1. الأمل الرياضي

A . تعريف

الأمل الرياضي (espérance mathématique) E{ X} للمتغيّرة العشوائية X هو المعدّل الوسطي الحسابي للقيم الممكنة مرجّحاً بواسطة الاحتمالات المناسبة .

حالة المتغيّرات المنفصلة

لنفترض أنّ p_i هو احتمال أن تأخذ المتغيّرة العشوائية X القيمة x_i :

$$E\{X\} = \sum_i p_i x_i .$$

---

(1) الرمز $\sum$ يعني مجموع ، ويُقرأ « سيغما » . $\sum_i x_i$ يعني مجموع القيم x_i..

إذا كانت مجموعة القيم الممكنة لا متناهية ولا يمكن تعداد عناصرها ، قد لا تتّجه السلسلة نحو حدّ معيّن . الأمل الرياضي يساوي مجموع هذه السلسلة على شرط أن تتّجه مطلقاً نحو حدّ معيّن :

$$E(X) = \lim_{n \to +\infty} \lim_{m \to +\infty} \sum_{i=-m}^{i=n} p_i x_i$$

وفي الحالة المعاكسة نقول أنّه لا وجود للأمل الرياضي .

مثل 1 . X هي عدد النقاط الحاصلة على حجر زهر .

$$E(X) = \sum_{x=1}^{6} \frac{1}{6} \cdot x = \frac{1}{6} \sum_{x=1}^{6} x = \frac{1}{6} \cdot \frac{6.7}{2} = 3{,}5 \quad (^1) .$$

مثل 2 . وعاء يحتوي كرات بيضاء بنسبة p وكرات حمراء بنسبة q $(q=1-p)$ وX هي متغيّرة برنولي العشوائية التي سبق أن حدّدناها ص 36 :

$$E(X) = p \times 1 + q \times 0 = p .$$

مثل 3. X هي عدد الرميات المتتالية لقطعة نقود والضرورية قبل الحصول على الوجه pile للمرّة الأولى . لقد رأينا (ص40) أنّ :

$$P(X = x) = \frac{1}{2^x} .$$

إذن :

$$E(X) = \sum_{x=1}^{x} \frac{x}{2^x} ,$$

أي :

$$\begin{array}{l} \frac{1}{2} \\ + \frac{1}{4} + \frac{1}{4} \\ + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ \cdots\cdots\cdots\cdots \\ + \frac{1}{2^x} + \frac{1}{2^x} + \frac{1}{2^x} + \cdots + \frac{1}{2^x} \end{array}$$

,

---

(1) المجموع S للـ n عدداً صحيحاً الأول يساوي $n(n+1)/2$ في الواقع :

$$\begin{array}{l} S = 1 + 2 + \cdots + (n-1) + n \\ \underline{S = n + (n-1) + \cdots + 2 + 1} \\ 2S = (n+1) + (n+1) + \cdots + (n+1) + (n+1) = n(n+1) \end{array}$$

$$E\{X\} = 1 + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^{n-1}} + \cdots$$

$$E\{X\} = 2,$$

..وذلك بجمعنا تباعاً المتواليات الهندسيّة ذات الأساس $\frac{1}{2}$ التي تؤلّف هذه السلسلة .

حالة المتغيّرات المتواصلة

لنفترض أنّ f(x) هي كثافة احتمال المتغيّرة العشوائية X عند النقطة x :

$$E\{X\} = \int_a^b xf(x)\,dx,$$

حيث a وb هما طرفا فسحة تحديد المتغيّرة X .

إذا كانت مجموعة التحديد ذات طول غير متناه ، فإنّ الأمل الرياضي هو غير محدّد إلّا إذا كان التكامل يتّجه مطلقاً نحو حدّ معيّن :

حدّ حدّ

$$E\{X\} = \lim_{a\to+\infty} \lim_{b\to+\infty} \int_{-a}^{b} xf(x)\,dx.$$

مثل 1 . لنفترض أنّ X هي المتغيّرة العشوائية المتواصلة الثابتة محدّدة على القطعة (0, 10) .

كثافة احتمال هذه المتغيّرة تساوي :

$$f(x) = 1/10.$$

في الواقع :

$$\int_0^{10} f(x)\,dx = \int_0^{10} \frac{dx}{10} = 1.$$

وأملها الرياضي هو :

$$E\{X\} = \int_0^{10} x\frac{dx}{10} = 5$$

مثل 2 . لنحدّد المتغيّرة العشوائية المسمّاة متغيّرة كوشي (Cauchy) على الفسحة

: بواسطة الكثافة $(-\infty, +\infty)$

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

$$E\{X\} = \int_a^b xf(x)\,dx = \int_a^b \frac{x\,dx}{\pi(1 + x^2)} = \frac{1}{2\pi}\ln\frac{1 + b^2}{1 + a^2}.$$

(lnيعني اللوغاريتم النبيري) .

عندما يميل a وb كلّ على حدة نحو اللانهاية (∞) ، فإنّ حد التكامل هو اللانهاية : إنّه لا يتّجه مطلقاً نحو حدّ معيّن ولا وجود للأمل الرياضي .

إلاّ أنّنا نشير إلى أنّه إذا مال a وb معاً نحو اللانهاية ، فإنّ التكامل يتّجه بفضل التوازن نحو الحدّ صفر (0) .

نشير هنا إلى صلة القرابة المتينة الموجودة بين تعريف الأمل الرياضي لمتغيّرة عشوائية وتعريف المعدّل الوسطي الحسابي لمتغيّرة إحصائية . في الحالة الأولى ، معايلات الترجيح هي الاحتمالات ؛ وفي الثانية ، الترددات الملحوظة .

للأمل الرياضي خصائص شبيهة بخصائص المعدّل الوسطي الحسابي .

B . خصائص الأمل الرياضي

1 . a وb هما ثابتان وX متغيّرة عشوائية :

$$E\{aX + b\} = aE\{X\} + b \qquad (1)$$

في الواقع ، في حالة المتغيّرة المنفصلة :

$$E\{aX + b\} = \sum_i p_i(ax_i + b)$$
$$= a\sum_i x_i p_i + b\sum_i p_i = aE\{X\} + b,$$

لأنّ تعريف الأمل الرياضي يعطي : $E\{X\} = \sum_i x_i p_i$ ولأنّ $\sum_i p_i = 1.$

كذلك في حالة المتغيّرة المتواصلة :

$$E\{aX + b\} = \int_{-\infty}^{+\infty} (ax + b)f(x)\,dx$$
$$= a\int_{-\infty}^{+\infty} xf(x)\,dx + b\int_{-\infty}^{+\infty} f(x)\,dx$$
$$= aE\{X\} + b.$$

هذه الخاصّة تعادل الخاصّة التي سمحت لنا باختزال حساب المعدّل الوسطي

الحسابي لمتغيّرة إحصائية عن طريق إبدال المتغيّرة (أنظر كتاب « الإحصاء الوصفي » ، الفصل V ، القسم I ، الفقرة 3.B) . ففي الواقع ، إذا أخذنا المتغيّرة المساعدة x'_i محدّدة بواسطة إبدال المتغيّرة التالي :

$$x_i = ax'_i + x_0 .$$

يوجد عندئذٍ بين المعدّلين الوسطيّين x وَ x' نفس العلاقة الخطّية الموجودة بين المتغيّرتين :

$$\bar{x} = a\bar{x}' + x_0 .$$

2 . لنفترض أنّ X وY هما متغيّرتان عشوائيتان :

$$E\{X + Y\} = E\{X\} + E\{Y\} .$$

الأمل الرياضي لحاصل جمع متغيّرتين عشوائيتين يساوي حاصل جمع الأملين الرياضيين لكلّ منهما .

حالة المتغيّرات المنفصلة

لنفترض أنّ p_{ij} هو احتمال أن تأخذ X القيمة x_i وY القيمة y_j :

$$p_{ij} = P\{X = x_i ; Y = y_j\}$$

$$\begin{aligned} E\{X + Y\} &= \sum_i \sum_j (x_i + y_j) p_{ij} \\ &= \sum_i \sum_j x_i p_{ij} + \sum_i \sum_j y_j p_{ij} \\ &= \sum_i x_i \sum_j p_{ij} + \sum_j y_j \sum_i p_{ij} \\ &= \sum_i x_i p_{i.} + \sum_j y_j p_{.j} . \end{aligned}$$

حيث الاحتمالات $p_{i.}$ تمثّل قانون احتمال X الهامشي والاحتمالات $p_{.j}$ تمثّل قانون احتمال Y الهامشي . إذن :

$$E\{X + Y\} = E\{X\} + E\{Y\} ,$$

لأنّ تعريف الأمل الرياضي يعطي :

$$E\{X\} = \sum_i x_i p_{i.} \quad \text{و} \quad E\{Y\} = \sum_j y_j p_{.j} .$$

حالة المتغيّرات المتواصلة

لنفترض أنّ $f(x, y)$ هي كثافة احتمال الزوج (X, Y) المكوّن من المتغيّرتين العشوائيتين X وY :

$$f(x, y)\,dx\,dy = P\{x \leqslant X < x + dx\,;\, y \leqslant Y < y + dy\},$$

$$E\{X+Y\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x+y)f(x, y)\,dx\,dy$$
$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}xf(x, y)\,dx\,dy + \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}yf(x, y)\,dx\,dy$$
$$= \int_{-\infty}^{+\infty}x\,dx\int_{-\infty}^{+\infty}f(x, y)\,dy + \int_{-\infty}^{+\infty}y\,dy\int_{-\infty}^{+\infty}f(x, y)\,dx$$
$$= \int_{-\infty}^{+\infty}xf(x)\,dx + \int_{-\infty}^{+\infty}yf(y)\,dy\,.$$

$\int_{-\infty}^{+\infty}f(x, y) = f(x)$ هي كثافة X الهامشية و $f(y) = \int_{-\infty}^{+\infty}f(x, y)\,dx$ هي كثافة Y الهامشية .

إذن :

$$E\{X+Y\} = E\{X\} + E\{Y\}$$

لأنّ تعريف الأمل الرياضي يعطي :

$$E\{X\} = \int_{-\infty}^{+\infty}xf(x)\,dx \qquad و \qquad E\{Y\} = \int_{-\infty}^{+\infty}yf(y)\,dy\,.$$

الخاصّتان السابقتان يجعلان من الأمل الرياضي مؤثراً خطياً ، وسوف تفيدانا لاحقاً (الفصل II وIII) لدراسة القانون ذي الحدّين وقانون توزيع المعدّل الوسطي لعيّنة ما .

تطبيق 1 : الأمل الرياضي للفرق بين متغيّرتين عشوائيتين :

إذا ضربنا في القاعدة (2) Y بـ 1 - ، نحصل على :

$$E\{X + (-Y)\} = E\{X\} + E\{-Y\},$$

إذن :

$$E\{X - Y\} = E\{X\} - E\{Y\},$$

بفضل الخاصّة 1 .

تطبيق 2 . الأمل الرياضي لمعدّل متغيرات عشوائية الوسطي .

لنفترض أنّ X_1 ، X_2 ، . . . . ، X_n هي n متغيّرة عشوائية تتبع قانون احتمال معيّن ذا أمل يساوي m . معدّلها الوسطي :

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

هو بدوره متغيّرة عشوائية، لنحسب أمله الرياضي :

$$E\{\bar{X}\} = E\left\{\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right\} = \frac{1}{n} E\{X_1 + X_2 + \cdots + X_n\},$$

بفضل الخاصّة 1 ، و

$$E\{\bar{X}\} = \frac{1}{n}[E\{X_1\} + E\{X_2\} + \cdots + E\{X_n\}],$$

بفضل الخاصّة 2 .

ولأنّ : $E\{X_1\} = E\{X_2\} = \cdots = E\{X_n\} = m,$

نستنتج أنّ : $E\{\bar{X}\} = m.$

3. لنفترض X وY متغيّرتين عشوائيتين مستقلّتين ، إذن :

$$E\{X.Y\} = E\{X\}.E\{Y\}.$$

الأمل الرياضي لحاصل ضرب متغيّرتين عشوائيّتين مستقلّتين يساوي حاصل ضرب الأملين الرياضيين لكلّ منهما .

حالة المتغيّرات المنفصلة

لنفترض أن p_{ij} هو احتمال أن تأخذ X القيمة x_i وY القيمة y_j :

$$E\{X.Y\} = \sum_i \sum_j p_{ij} x_i y_j.$$

وبما أنّ المتغيّرتين X وY مستقلّتان :

$$p_{ij} = p_{i.} \times p_{.j}.$$

بالتالي :

$$E\{X.Y\} = \sum_i \sum_j p_{i.} x_i \times p_{.j} y_j = \sum_i p_{i.} x_i \times \sum_j p_{.j} y_j$$

إذن :

$$E\{X.Y\} = E\{X\} \times E\{Y\}.$$

حالة المتغيّرات المتواصلة

لنفترض f(x, y) كثافة احتمال الزوج (X, Y) المكوّن من المتغيّرتين العشوائيتين X وY :

$$E\{X.Y\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xyf(x, y)\,dx\,dy .$$

وبما أنّ المتغيّرتين X وY مستقلّتان :

$$f(x, y) = f(x).f(y) .$$

بالتالي :

$$E\{X.Y\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xf(x)\,dx \times yf(y)\,dy$$

$$= \int_{-\infty}^{+\infty} xf(x)\,dx \times \int_{-\infty}^{+\infty} yf(y)\,dy$$

إذن :

$$E\{X.Y\} = E\{X\} \times E\{Y\} .$$

2 . التباين

A . تعريف

التباين (variance) V { X } للمتغيّرة العشوائية X هو الأمل الرياضي لمربّعات الفوارق بين قيم المتغيّرة وأملها الرياضي :

$$V\{X\} = E\{(X - E\{X\})^2\}$$

حالة المتغيّرات المنفصلة

$$V\{X\} = \sum_i p_i(x_i - E\{X\})^2 .$$

حالة المتغيّرات المتواصلة

$$V\{X\} = \int_{-\infty}^{+\infty} (x - E\{X\})^2 f(x)\,dx .$$

الانحراف النموذجي σ_x (سيغما إكس) هو الجذر التربيعي للتباين :

$$\sigma_x = \sqrt{V\{X\}} .$$

ولهذا السبب يُسمّى التباين مربّع الانحراف النموذجي .

مثل 1 . وعاء يحتوي كرات بيضاء بنسبة p وكرات حمراء بنسبة q‏(q=1−p) ، X هي متغيّرة برنولي العشوائية المحدّدة في القسم IV ، ص 36 .

أملها الرياضي المحسوب ص ــ هو :

$$E\{X\} = p.$$

بالتالي :

$$V\{X\} = p(1-p)^2 + q(0-p)^2 = pq^2 + qp^2$$
$$= pq(p+q) = pq, \quad \text{لأنّ} \quad p+q=1$$

مثل 2 . X هي عدد النقاط الحاصلة على حجر زهر .

سبق أن حسبنا أملها الرياضي ص 53 :

$$E\{X\} = 3{,}5.$$

بالتالي :

$$V\{X\} = \frac{1}{6}[(1-3{,}5)^2+(2-3{,}5)^2+(3-3{,}5)^2+(4-3{,}5)^2+(5-3{,}5)^2+(6-3{,}5)^2]$$

$$= \frac{1}{3}[(2{,}5)^2+(1{,}5)^2+(0{,}5)^2] = \frac{8{,}75}{3} = \frac{35}{12} \approx 2{,}92.$$

نشير هنا أيضاً إلى التقارب الحاصل بين التباين ، أو الانحراف النموذجي ، لمتغيّرة عشوائية والتباين ، أو الانحراف النموذجي ، لمتغيّرة إحصائية ، ولكليهما الخصائص نفسها .

B . خصائص التباين

1 . لنفترض أنّ a وb هما ثابتان وX متغيّرة عشوائية ، إذن :

$$V\{aX+b\} = a^2 V\{X\}. \quad (1)$$

في الواقع :

$$V\{aX+b\} = E\{(aX+b-E\{aX+b\})^2\}.$$

ولكن مع الأخذ بخصائص الأمل الرياضي :

$$E\{aX+b\} = aE\{X\}+b,$$

إذن :

$$V\{aX+b\} = E\{a^2(X-E\{X\})^2\}$$
$$= a^2 E\{(X-E\{X\})^2\} = a^2 V\{X\}.$$

هذه الخاصّة تعادل الخاصة التي سمحت لنا باختزال حساب التباين لمتغيّرة

إحصائية عن طريق إبدال المتغيّرة (أنظر كتاب « الإحصاء الوصفي » ، الفصل V ، القسم II ، الفقرة 4.B) . فلنأخذ في الحقيقة المتغيّرة المساعدة x_i' المحدّدة بواسطة إبدال المتغيّرة التالي :

$$x_i = ax_i' + x_0 .$$

يوجد بين التباينين V { X } و V { X' } العلاقة التالية :

$$V\{x\} = a^2 V\{x'\} .$$

2 . لنفترض أنّ X وY هما متغيّرتان عشوائيتان مستقلّتان ، إذن :

$$V\{X + Y\} = V\{X\} + V\{Y\} .$$

إنّ تباين حاصل جمع متغيّرتين عشوائيتين مستقلّتين يساوي حاصل جمع التباينين لكلّ منهما .

في الحقيقة ، إنطلاقاً من تعريف التباين :

$$V\{X + Y\} = E\{[X + Y - E\{X + Y\}]^2\}$$

وتبعاً لخصائص الأمل الرياضي :

$$V\{X+Y\}=E\{[(X-E\{X\})+(Y-E\{Y\})]^2\}$$
$$=E\{(X-E\{X\})^2\}+E\{(Y-E\{Y\})^2\}+2E\{(X-E\{X\})(Y-E\{Y\})\}.$$

إلّا أنّنا سوف نبرهن في الفقرة التالية أنّ العبارة :

$$E\{(X - E\{X\})(Y - E\{Y\})\},$$

التي نسمّيها تغاير المتغيّرتين X وY ، تساوي صفراً عندما تكون المتغيّرتان X وY مستقلّتين .

بالتالي :

$$V\{X + Y\} = V\{X\} + V\{Y\} .$$

هذه الخصائص ، مثل خصائص الأمل الرياضي ، سوف تفيدنا عند دراستنا للقانون ذي الحدّين ولقانون توزيع المعدّل الوسطي لعيّنة ما .

تطبيق 1 : تباين الفارق بين متغيّرتين عشوائيتين مستقلّتين .
إذا ضربنا في القاعدة (2) Y بـ 1− ، نحصل على :

$$V\{X+(-Y)\}=V\{X\}+V\{-Y\}=V\{X\}+V\{Y\}$$

إذن :

$$V\{X-Y\}=V\{X\}+V\{Y\}$$

وذلك بفضل الخاصة 1 .

تطبيق 2 : تباين المعدّل الوسطي لمتغيّرات عشوائية مستقلّة .

لنفترض أن :

$$\overline{X}=\frac{X_1+X_2+\cdots+X_n}{n}$$

هي المعدّل الوسطي لـ n متغيّرة عشوائية مستقلّة تتبع جميعها قانون احتمال معيّن ذي أمل رياضي m وتباين σ^2 . انطلاقاً من تعريف X :

$$V\{\overline{X}\}=V\left\{\frac{1}{n}(X_1+X_2+\cdots+X_n)\right\}=\frac{1}{n^2}V\{X_1+X_2+\cdots+X_n\}$$

بفضل الخاصّة 1 ؛ و

$$V\{\overline{X}\}=\frac{1}{n^2}[V\{X_1\}+V\{X_2\}+\cdots+V\{X_n\}].$$

بفضل الخاصّة 2 .

وبما أنّ : $V\{X_1\}=V\{X_2\}=\cdots=V\{X_n\}=\sigma^2$

إذن $V\{\overline{X}\}=\sigma^2/n$

3 ـ تغاير متغيّرتين عشوائيتين

لنفترض أنّ X وY هما متغيّرتان عشوائيتان ، نعرّف تغاير (Covariance) X وY كالتالي :

$$\text{cov}\{X,Y\}=E\{(X-E\{X\})(Y-E\{Y\})\}$$

خاصة . إنّ تغاير متغيّرتين عشوائيتين مستقلّتين يساوي صفراً .

لنأخذ المتغيّرتين المركزتين :

$$X'=X-E\{X\}\qquad Y'=Y-E\{Y\}$$

تغاير X وY يُكتب :

$$\text{cov}\{X,Y\}=E\{X'Y'\}$$

وبما أنّ X وY مستقلّتان ، فإنّ X' وY' مستقلّتان أيضاً . بالتالي ، وبفضل

خصائص الأمل الرياضي لحاصل ضرب متغيّرتين مستقلّتين :

$$\text{cov}\{X, Y\} = E\{X' Y'\} = E\{X'\}.E\{Y'\}.$$

لكن تعريف المتغيّرات المركزة يعطينا :

$$E\{X'\} = E\{X - E\{X\}\} = 0$$
$$E\{Y'\} = E\{Y - E\{Y\}\} = 0$$

إذن :

$$\text{cov}\{X, Y\} = 0.$$

4 . العزم

العزم (moment) من الدرجة k للمتغيّرة العشوائية X هو الأمل الرياضي للمتغيّرة X^k :

$$m_k = E\{X^k\}.$$

العزم من الدرجة 1 : $m_1 = E\{X\}.$

العزم من الدرجة 2 : $m_2 = E\{X^2\}.$

الخ .

بالإمكان بسط هذه الفكرة إلى زوج من المتغيّرات العشوائية (X, Y) . العزم من الدرجة (r, s) هو :

$$m_{rs} = E\{X^r Y^s\}$$

$$m_{10} = E\{X\}.$$
$$m_{01} = E\{Y\}.$$
$$m_{20} = E\{X^2\}.$$
$$m_{02} = E\{Y^2\}.$$
$$m_{11} = E\{XY\}.$$

الخ .

التعبير عن التباين بواسطة العزم

إن تعريف التباين يعطينا :

$$V\{X\} = E\{(X - E\{X\})^2\}$$
$$= E\{X^2 - 2XE\{X\} + E\{X\}^2\}.$$

بفضل خصائص الأمل الرياضي :

$$V\{X\} = E\{X^2\} - 2E\{X\}^2 + E\{X\}^2$$
$$= E\{X^2\} - E\{X\}^2$$
$$V\{X\} = m_2 - m_1^2 .$$

هذه العبارة تطابق القاعدة الموسّعة التي استعملناها لإنجاز حساب التباين لمتغيّرة إحصائيّة (أنظر كتاب « الإحصاء الوصفي » ، الفصل V ، القسم II ، الفقرة 4.B) وهي تسمح ، بالطريقة ذاتها ، باختزال حساب تباين متغيّرة عشوائية .

مثلاً . X هي عدد النقاط الحاصلة على حجز زهر .

$$m_2 = E\{X^2\} = \frac{1}{6}\sum_{x=1}^{6} x^2 = \frac{1}{6}\frac{6.7.13}{6} = \frac{91}{6} \quad (^1).$$

إذن :
$$m_1 = E\{X\} = 3{,}5 ,$$

$$V\{X\} = m_2 - m_1^2 = \frac{91}{6} - \frac{49}{4} = \frac{35}{12}.$$

يمكننا ، بالطريقة ذاتها ، أن نعبّر بواسطة العزم عن التغاير بين زوج من المتغيّرات العشوائية (X, Y) :

$$\text{cov}(XY) = E\{(X - E\{X\})(Y - E\{Y\})\}$$
$$= E\{XY - XE\{Y\} - YE\{X\} + E\{X\}E\{Y\}\}$$
$$= E\{XY\} - E\{X\}E\{Y\} - E\{Y\}E\{X\} + E\{X\}E\{Y\}$$
$$= E\{XY\} - E\{X\}E\{Y\},$$
$$\text{cov}(XY) = m_{11} - m_{10}m_{01} .$$

---

(1) إنّ حاصل جمع مربّعات الـ n عدداً صحيحاً الأولى يساوي $\frac{n(n+1)(2n+1)}{6}$ في الواقع :

$$(1+1)^3 = 1 + 3.1 + 3.1^2 + 1^3$$
$$(1+2)^3 = 1 + 3.2 + 3.2^2 + 2^3$$
$$(1+3)^3 = 1 + 3.3 + 3.3^2 + 3^3$$
$$\cdots\cdots\cdots\cdots$$
$$(1+n)^3 = 1 + 3.n + 3.n^2 + n^3$$
$$(1+n)^3 = n + 3\sum_{i=1}^{n} i + 3\sum_{i=1}^{n} i^2 + 1$$

إذن ، $\sum_{i=1}^{n} i = \frac{n(n+1)}{2}$:

$$3\sum_{i=1}^{n} i^2 = (1+n)^3 - n - 3\frac{n(n+1)}{2} - 1$$
$$\sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}$$

الفصل الثاني

قوانين التوزيع الإحصائي
النماذج المنفصلة

إنّ معظم الظواهر الإحصائية يمكن أن تُشرح بواسطة عدد صغير من النماذج الاحتمالية أو قوانين الاحتمال . وعندما يكون هذا التمثيل ممكناً فإنّه يعطي وصفاً للظاهرة أغنى من مجرّد حساب المميّزات ذات الميل المركزي ومميّزات التفرّق . فهو يسمح مثلاً بحساب احتمال بعض الحوادث ويحدّد بالتالي بشكل ما التمثيل الذي يمكن تصوّره لمستقبل هذه الظاهرة .

ينبغي إذن أن نتعرّف إلى النماذج الاحتمالية الأكثر انتشاراً بشكل يسمح لنا بالبحث في هذه القائمة عن النموذج المناسب لوصف ظاهرة عشوائية معيّنة .

في كلّ الأحوال ، الإجراء يتمّ كالتالي :

- تعطينا ملاحظة الظاهرة توزيعاً اختبارياً أو تجريبياً .

- تحليل هذا التوزيع التجريبي - أي فحص التمثيل البياني وحساب المميّزات ذات الميل المركزي ومميّزات التفرّق - يعطي فكرة أولى عن طبيعة الظاهرة الملحوظة . عند رؤية هذه النتائج الأولى ، نختار بين مختلف قوانين التوزيع النظري قانوناً نراه مناسباً ، وهذا يعني أن نختار شكل « القالب » الذي نستطيع أن « نصبّ » فيه الظاهرة . يجب إذن ، إنطلاقاً من السلسلة التجريبية ، تقدير متغيّرات هذا القانون الوسيطية ، وهذا يعني اختيار « القالب » ذي الحجم المناسب .

- بالطبع لا يُعتبر استبدال التوزيع التجريبي بالقانون النظري صحيحاً إلاّ إذا كانت القيم الملحوظة قريبة بشكل كاف من القيم النظرية الناتجة عن النموذج : يجب اختيار

كون الوصف الذي يعطيه القانون النظري للظاهرة مقبولاً ، بعبارة أخرى كون الفوارق الملحوظة بين الترددات التجريبية والترددات النظرية عائدة إلى عامل الصدفة .

القسم I

القانون ذو الحدّين

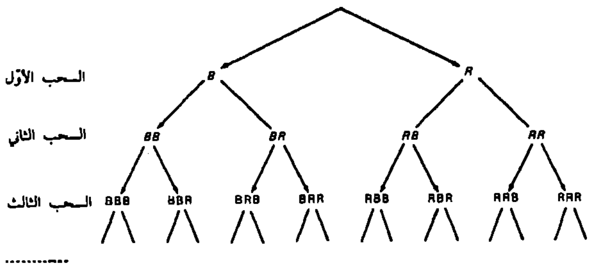
1 . تعريف .ـ 2 . شروط التطبيق .ـ 3 . المتغيّرة ذات الحدّين كمجموع متغيّرات برنولي عشوائية مستقلة .ـ 4 . المقاييس : A . المنوال ؛ B . الأمل الرياضي ؛ C . التباين .ـ 5 . قانون احتمال ومقاييس تردّد متغيّرة ذات حدّين .ـ 6 . تحديد الاحتمالات عملياً .ـ 7 . تسوية قانون ذي حدّين مع توزيع إحصائي ملحوظ .

نصادف القانون ذا الحدّين كلّ مرة نقع فيها على خيارين يبقى احتمالاهما ثابتين على مرور سلسلة من التجارب : صبي أو بنت ، موت أو حياة ، قبول أو رفض قطع تُصنع بالجملة ، الخ . وتأتي أهمّية هذا القانون ، بصورة خاصة ، من كونه يُطبّق على سحب عيّنة عشوائية وعلى تفسير النتائج المنبثقة عن هذه الطريقة .

1 . تعريف

لنأخذ وعاء يحتوي N كرة في فئتين :

ـ كرات بيضاء B بنسبة p ،



الشكل 21 . رسم بياني لشجرة الحوادث الممكنة

ـ كرات حمراء R بنسبة $q = 1 - p$

نجري n سحباً متتالياً لكرة واحدة ، مع ردّها كلّ مرّة إلى الوعاء قبل السحب التالي . نعرّف المتغيّرة العشوائية ذات الحدّين X كعدد الكرات البيضاء التي نحصل عليها خلال السحوبات الـ n . بإمكانها أن تأخذ القيم التالية : 0، 1 ، 2، . . . ، k ، . . . ، n . إنّها متغيّرة منفصلة .

قانون الاحتمال

يمكننا الحصول على مختلف الحوادث الممكنة تبعاً لرسم شجرة.(الشكل 21) :
عند السحب الأوّل ، قد نحصل على كرة بيضاء B أو على كرة حمراء R ؛ عند السحب الثاني ، سواء كانت الكرة الملحوظة أوّلاً بيضاء أو حمراء ، فإنّنا قد نحصل من جديد إمّا على كرة بيضاء إمّا على كرة حمراء ، الخ . عند كل سحب إذن هناك خياران لا ثالث لهما وعدد الحوادث الممكنة خلال n سحباً يساوي 2^n .

تسمح طريقة المعالجة هذه بتحديد مختلف الإمكانيات المحتملة وقانون احتمال المتغيّرة ذات الحدّين X المناسبة لعدد n من السحوبات التالية :

الحدث النموذجي	المتغيّرة العشوائية X	الاحتمال $P(X)$
	السحب الأوّل : $n = 1$	
B	1	p
R	0	q
		المجموع : 1
	السحب الثاني : $n = 2$	
BB	2	p^2
BR RB	1	$2pq$
RR	0	q^2
		المجموع : 1
	السحب الثالث : $n = 3$	
BBB	3	p^3
BBR BRB RBB	2	$3p^2q$

BRR RBR RRB	1	$3\,pq^2$
RRR	0	q^3
		المجموع : 1

عند السحب الثالث مثلاً ، تأخذ المتغيّرة X القيمة 2 لكلّ من الحوادث النموذجية التالية :

BBR, BRB, RBB

يساوي احتمال كلّ من هذه الحوادث p^2q (قاعدة الاحتمالات المركّبة) ، أمّا احتمال أن تكون المتغيّرة X تساوي 2 ، وهي قيمة تطابق تحقيق حدث أو آخر من الحوادث الثلاثة النموذجية ، فيساوي $3p^2q$ (قاعدة الاحتمالات الكلّية) :

$$P_2 = P\{X = 2\} = 3\,p^2\,q\,.$$

عند السحب رقم n ، تأخذ المتغيّرة X القيمة x لكلّ حدث نموذجي يطابق ظهور x كرة بيضاء . ويساوي احتمال كلّ من هذه الحوادث $p^x q^{n-x}$ (قاعدة الاحتمالات المركّبة) ، وهناك C_n^x حدثاً من هذا النوع : فإنّ احتمال أن تأخذ المتغيّرة X القيمة x المطابقة لتحقيق حدث أو آخر من هذه الحوادث الـ C_n^x النموذجية يساوي $C_n^x p^x q^{n-x}$ (قاعدة الاحتمالات الكلّية) :

$$P_x = P\{X = x\} = C_n^x\,p^x\,q^{n-x}$$

وهكذا تظهر الاحتمالات كعناصر توسيع ذي الحدّين $(p+q)^n$ ، حيث n هو عدد السحوبات المنجزة :

السحب الأوّل : $p + q$

السحب الثاني : $(p+q)^2 = p^2 + 2\,pq + q^2$

السحب الثالث : $(p+q)^3 = p^3 + 3\,p^2\,q + 3\,pq^2 + q^3$

....

السحب رقم n :

$$(p+q)^n = p^n + C_n^{n-1}\,p^{n-1}\,q + C_n^{n-2}\,p^{n-2}\,q^2 + \cdots + C_n^x\,p^x\,q^{n-x}$$
$$+ C_n^1\,pq^{n-1} + q^n$$

$$= p^n + np^{n-1} q + \frac{n(n-1)}{2!} p^{n-2} q^2$$
$$+ \cdots + \frac{n(n-1)\ldots(n-x+1)}{x!} p^x q^{n-x} + \cdots + npq^{n-1} + q^n .$$

يمكننا التحقّق بهذه المناسبة ، مهما كان العددان n وp ، من كون مجموع كلّ الاحتمالات يساوي واحداً :

$$\sum_{x=0}^{n} P_x = \sum_{x=0}^{n} C_n^x p^x q^{n-x} = (p+q)^n = 1 ,$$ وذلك لأنّ $p + q = 1$

باختصار ، فإنّ القانون ذا الحدّين يتعلّق بمتغيّرين وسيطيّين (paramétres) :

n : وهو عدد السحوبات المتالية أو التجارب المستقلّة . ويمثّل ، في استقصاء بواسطة البحث الإحصائي ، مقدار العيّنة ؛

p : وهو احتمال تحقيق الحدث المدروس عند كلّ من السحوبات أو التجارب المستقلّة (نسبة الكرات البيضاء الموجودة في الوعاء) .

احتمال أن تأخذ المتغيّرة ذات الحدّين X القيمة x هو :

$$P\{X = x\} = C_n^x p^x q^{n-x} .$$

ونرمز إلى المتغيّرة X بواسطة :

$$X = \mathcal{B}(n, p),$$

للإشارة إلى أنّ المتغيّرة العشوائية X تتبع قانوناً ذا حدّين ومتغيّرين وسيطيّين n وp .

الشكل

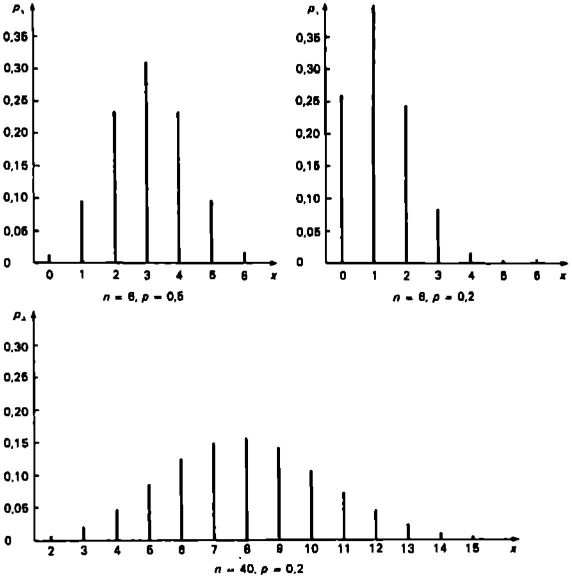
يكون التوزيع ذو الحدّين متناظراً (symétrique) عندما يكون q=p=0,5 ، ويكون غير متماثل في الحالة المعاكسة ، حيث يكبر اللاتماثل بمقدار ما يزداد الفارق بين p وq . إلاّ أنّه عندما يكون عدد الحالات الملحوظة كبيراً ، بشرط أن لا تكون p قريبة جدّاً من 0 أو من 1 ، فإن هذا التوزيع يميل إلى التناظر (الشكل 22) . في هذه الحالة ، سنرى لاحقاً أنّ التوزيع ذا الحدّين يقترب من التوزيع الطبيعي (المعتدل ، normal) .

2 . شروط التطبيق

إنّ مسألة الوعاء الذي نجري عليه n سحباً متتالياً مع ردّ الكرة المسحوبة هي

صورة : فالقانون ذو الحدّين يطبّق كلّ مرّة نقع فيها على خيارين A و$\bar{A}$ يبقى احتمالاهما ثابتين على مرور سلسلة من التجارب المستقلّة . يمكننا مثلاً تصوّر سياق صناعة بالجملة كسحب n عنصراً من المجتمع الإحصائي المتصوّر المكوّن من مجموعة القطع التي يمكن صنعها بواسطة الآلة . ويتضمّن هذا المجتمع الإحصائي المتصوّر نسبة ثابتة p من القطع التي لا تخضع لقواعد الصناعة ونسبة $q=1-p$ من القطع المقبولة . إذا كان بالإمكان تطبيق هذا النموذج ، فإنّ توزيع احتمال عدد القطع المعيبة هو قانون ذو حدّين .

ويطابق القانون ذو الحدّين بشكل خاص سياق سحب عيّنة عشوائية . لنفترض



الشكل 22 . شكل القانون ذي الحدّين

أنّا نبحث عن عدد الأشخاص الذين يستهلكون منتوجاً معيّناً واسع الانتشار . يمكننا تقسيم الشعب الى فئتين : الأشخاص الذين يستهلكون هذا المنتوج ، وعددهم N_1 ، والأشخاص الذين لا يستهلكونه ، وعددهم N_2 . تقوم طريقة الأبحاث الإحصائية على تعيين عيّنة من الأشخاص نسألهم ما إذا كانوا يستهلكون هذا المنتوج ، بعد سحبهم بالقرعة من ضمن الشعب ، وهذا نهج يعني سحب الأشخاص الذين يُسألون من وعاء (الشعب) يحتوي على فئتين (الأشخاص الذين يستهلكون المنتوج والذين لا يستهلكونه) .

إذا أجرينا السحوبات مع ردّ ما يُسحب فإنّ الاحتمال p أن نعيّن خلال واحد من السحوبات المتالية شخصاً يستهلك المنتوج يساوي :

$$p = \frac{N_1}{N_1 + N_2},$$

والاحتمال q أن نختار شخصاً لا يستهلك المنتوج يساوي :

$$q = \frac{N_2}{N_1 + N_2} = 1 - p.$$

ستسمح لنا إجراءات التقدير التي سنعرضها لاحقاً (أنظر الفصل VI) أن نستنتج انطلاقاً من عدد الأشخاص في العيّنة الذين يستهلكون المنتوج موضوع الدراسة ، عدد الأشخاص الذين يستهلكونه في المجتمع الإحصائي ، مع إشارة إلى مدى دقّة النتيجة التي نحصل عليها بهذه الطريقة . وتستند إجراءات التقدير (estimation) هذه إلى تمثيل سحب العيّنة بواسطة القانون ذي الحدّين .

في الواقع ، ، عندما نأخذ عيّنة ما فإنّنا نعمد إلى سحب مستنفد (tirage exhaustif) ـ ما يعني أنّنا لا نردّ الكرة الحاصلة إلى الوعاء بعد كلّ سحب ـ بشكل لا نعيّن معه نفس الفرد مرّتين . إنّ هذا النوع من سحب العيّنات يُمثّل ، على وجه الدّقة ، بواسطة القانون فوق الهندسي (hypergéométrique) ، الذي سنعرضه في الفقرة اللاحقة ، وليس بواسطة القانون ذي الحدّين . إلاّ أنّه عندما يكون مقدار المجتمع الإحصائي N كبيراً جدّاً بالنسبة لمقدار العيّنة n ، فإنّ الاحتمالين p وq يبقيان تقريباً ثابتين ويبقى القانون ذو الحدّين صالحاً .

3 . تأويل المتغيّرة ذات الحدّين كمجموع متغيّرات برنولي عشوائية مستقلّة

لنعد إلى مثل الوعاء الذي يحتوي :

ـ كرات بيضاء B بنسبة p ،
ـ كرات حمراء R بنسبة $q = 1 - p$
ولنجرِ n سحباً مع ردّ .

يمكننا عند كلّ سحب تحديد متغيّرة برنولي عشوائية ، مبيّنة للحدث : وهو الإشارة إلى سحب كرة بيضاء (أنظر ص 36) . وهذه المتغيّرة هي من ناحية أخرى شبيهة بالمتغيّرة ذات الحدّين المطابقة لتجربة واحدة

سوف ننسب متغيّرة برنولي X_i الى السحب ذي الرتبة i :

الاحتمال $P\{X_i\}$	المتغيّرة العشوائية X_i	الحدث النموذجي
p	1	B
q	0	R
المجموع : 1		

المتغيّرة ذات الحدّين X ، وهي عدد الكرات البيضاء الحاصلة خلال n سحباً ، تساوي مجموع n متغيّرة برنولي مستقلّة X_1 ، X_2 ، . . . ، X_n :

$$X = X_1 + X_2 + \ldots X_n$$

هذه المتغيّرات هي مستقلّة لأنّنا نعيد الكرة إلى الوعاء بعد كلّ سحب : إذن يبقى الاحتمالان p وq ثابتين ولا يتوقّفان على لون الكرات المأخوذة عند السحوبات السابقة بعكس الحالة التي نجري فيها السحوبات دون ردّ .

لنذكّر بمقاييس متغيّرة برنولي العشوائية ، أي الأمل الرياضي والتباين (أنظر ص 53 و60) .

الأمل الرياضي

$$E\{X_i\} = \sum_{x=0}^{1} xP\{X_i = x\} = 0 \times q + 1 \times p = p.$$

التباين

$$V\{X_i\} = \sum_{x=0}^{1} (x-p)^2 P\{X_i = x\} = (0-p)^2 \times q + (1-p)^2 \times p$$
$$= p^2 q + q^2 p = pq(p+q) = pq.$$

سوف يفيدنا هذا التأويل للمتغيّرة ذات الحدّين كمجموع متغيّرات برنولي

مستقلّة في حساب الأمل الرياضي والتباين للقانون ذي الحدّين .

4 . مقاييس القانون ذي الحدّين

A . المنوال

إنّ منوال القانون ذي الحدّين هو القيمة الصحيحة المحصورة بين np−q وnp+p .

البرهان : إنّ منوال توزيع احتمال معيّن هو قيمة المتغيّرة العشوائية صاحبة الاحتمال الأعلى : إنّها القيمة الأكثر احتمالاً .

بالتالي ، فإنّ منوال القانون ذي الحدّين هو العدد الصحيح x حيث :

$$P_{x-1} < P_x \qquad و \qquad P_x > P_{x+1} .$$

وهذا ما يمكننا كتابته أيضاً :

$$\frac{P_x}{P_{x-1}} > 1 \quad (1) \qquad و \qquad \frac{P_{x+1}}{P_x} < 1 \quad (2) .$$

لنحسب إذن نسبة الاحتمالين المنسوبين إلى قيمتين متتاليتين للمتغيّرة ذات الحدّين :

$$\frac{P_{x+1}}{P_x} = \frac{C_n^{x+1}\, p^{x+1}\, q^{n-x-1}}{C_n^x\, p^x\, q^{n-x}} = \frac{n!}{(x+1)!\,(n-x-1)!} \cdot \frac{x!\,(n-x)!}{n!} \cdot \frac{p}{q} = \frac{n-x}{x+1} \cdot \frac{p}{q} .$$

بالتالي ، يكتب التفاوتان (inégalités) (1) و(2) على الشكل :

$$\frac{P_{x+1}}{P_x} = \frac{n-x}{x+1} \cdot \frac{p}{q} < 1 \qquad (3)$$

$$\frac{P_x}{P_{x-1}} = \frac{n-x+1}{x} \cdot \frac{p}{q} > 1 \qquad (4)$$

وذلك بوضع x−1 مكان x في التفاوت الأوّل) .

نستنتج من (3) :

$$(n-x)\,p < (x+1)\,q ,$$
$$np - xp < x - xp + q ,$$
$$np - q < x .$$

ومن (4) :

$$(n-x+1)\,p > xq ,$$
$$np - xp + p > x - xp ,$$
$$np + p > x .$$

إذن :

$$np - q < x < np + p .$$

إذا كانت الكمّية np−q عدداً صحيحاً :

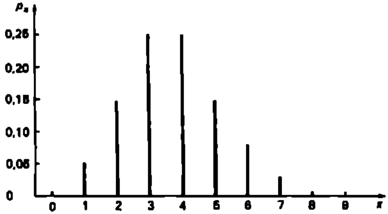
$np - q = i,$

$np + p = np - q + (p + q) = i + 1,$

np + p هو إذن العدد الصحيح اللذي يليها مباشرة . هناك إذن قيمتان منوال : np − q وnp + p .

مثلاً : n = 9 ، p = 0,4 . يوجد قيمتان منوال (الشكل 23) :

np − q = 3,6 − 0,6 = 3 و np + p = 3,6 + 0,4 = 4



الشكل 23 . قانون ذو حدّين بمتغيّرين وسيطين: n=9 وp=0,4 يوجد قيمتان منوال : 3 و4

B . الأمل الرياضي

يمكننا أن نعتبر المتغيّرة ذات الحدّين X ، المطابقة لـ n سحباً ، كمجموع n متغيّرة برنولي مستقلّة (أنظر ص 72) :

$$X = X_1 + X_2 + \cdots + X_n = \sum_{i=1}^{n} X_i .$$

أملها الرياضي :

$$E\{X\} = E\{X_1 + X_2 + \cdots + X_n\} = E\left\{\sum_{i=1}^{n} X_i\right\}$$

$$= E\{X_1\} + E\{X_2\} + \cdots + E\{X_n\} = \sum_{i=1}^{n} E\{X_i\} .$$

في الـواقع ، وبفضـل خصائص الأمـل الريـاضي ، فإنّ أمـل مجموع عـدد من المتغيّـرات العشوائية الرياضي يساوي مجموع الآمال الرياضية (أنـظر ص 56) .
وأمل متغيّـرة برنولي X_i الرياضي ، المحدّدة عند كلّ من السـحوبات هو :

$$E\{X_i\} = p.$$

بالتالي :

$$E\{X\} = \sum_{i=1}^{n} E\{X_i\} = np.$$

أمل التوزيع ذي الحدّين الرياضي (أو معدّله الوسطي) يساوي np .

C . التباين

إنّ تباين المتغيّـرة ذات الحدّين X هو :

$$V\{X\} = V\{X_1 + X_2 + \cdots + X_n\} = V\left\{\sum_{i=1}^{n} \right\}$$

$$= V\{X_1\} + V\{X_2\} + \cdots + V\{X_n\} = \sum_{i=1}^{n} V\{X_i\}.$$

لأنّه في الواقع ، وبفضل خصائص التباين ، فإنّ تباين مجموع عدد من المتغيّـرات العشوائية المستقلّـة يساوي مجموع التباينات (أنظر ص 61) . وتباين متغيّـرة برنولي X_i ، المحدّدة عند كلّ من السحوبات هو :

$$V\{X_i\} = pq.$$

بالتالي :

$$V\{X\} = \sum_{i=1}^{n} V\{X_i\} = npq.$$

ويساوي انحراف التوزيع ذي الحدّين النموذجي $\sqrt{npq}$

5. قانون احتمال ومقاييس التردّد ذي الحدّين

لنفترض أنّ X هي متغيّـرة عشوائية ذات حدّين بمتغيّـرين وسيطيّين n و p :

$$X = \mathcal{B}(n, p).$$

لنركّـز اهتمامنا الآن ، ليس على الكرات البيضاء X المسحوبة أثناء الـ n تجربة مستقلّـة ، بل على التردّد (fréquence) f_x لهذا الحدث : $f_X = \frac{X}{n}.$

تمثّل هذه المتغيّـرة نسبة التجارب حيث تمّ تحقيق الحدث « الحصـول على كـرة بيضاء » .

يُستنتج قانون توزيع f_X مباشرة من قانون توزيع X :

$$P\left\{f_X = \frac{x}{n}\right\} = P\{X = x\} = C_n^x p^x q^{n-x}.$$

مثلاً ، بالنسبة لسحب ثلاث كرات من وعاء :

الحدث النموذجي	المتغيّرة العشوائية X	التردّد ذو الحدّين $f_X = \frac{X}{n}$	الاحتمال
BBB	3	1	p^3
BBR BRB RBB	2	2/3	$3p^2q$
BRR RBR RRB	1	1/3	$3pq^2$
RRR	0	0	q^3
			المجموع: 1

الأمل الرياضي

بوسعنا أن نكتب :

$$E\{f_X\} = E\left\{\frac{X}{n}\right\} = \frac{1}{n}E\{X\},$$

وذلك تبعاً لخاصّة الأمل الرياضي التالية :

$$E\{aX\} = aE\{X\}$$ (أنظر ص 55).

وبما أنّ أمل المتغيّرة ذات الحدّين الرياضي يساوي np :

$$E\{f_X\} = p.$$

أمل التردّد ذي الحدّين $f_X = \frac{X}{n}$ الرياضي يساوي p ، وهو احتمال تحقيق الحدث موضوع الدراسة (مثلاً ، الحصول على كرة بيضاء) عند كلّ من السحوبات .

التباين

يمكننا كذلك الكتابة :

$$V\{f_X\} = V\left\{\frac{X}{n}\right\} = \frac{1}{n^2}V\{X\},$$

وذلك تبعاً لخاصّة التباين التالية :

$V\{aX\} = a^2 V\{X\}$ (أنظر ص 61) .

(أنظر ص 61)

وبما أنّ تباين المتغيّرة ذات الحدّين يساوي npq :

$$V\{f_X\} = \frac{pq}{n}.$$

ويساوي انحراف التردّد ذي الحدّين $f_X = X/n$ النموذجي $\sqrt{pq/n}$.

6 . حساب الاحتمالات العملي . جداول القانون ذي الحدّين

إنّ حساب القيمة العددية للاحتمال المنسوب إلى كلّ قيمة لِـ X :

$$P\{X = x\} = C_n^x p^x q^{n-x}.$$

يصبح مملّاً عندما يكبر العدد n نسبياً .

مثلاً . نرمي بحجر زهر 5 مرّات ونهتمّ بالمتغيّرة ذات الحدّين X : عدد المرّات التي نحصل فيها على الرقم 1 .

متغيّرا هذا القانون ذي الحدّين الوسيطيان هما n=5 وp=1/6 . لنحسب مثلاً احتمال أن يكون العدد X ، عدد المرات التي نحصل فيها على 1 ، يساوي 3 هو :

$$P_3 = P\{X = 3\} = C_5^3 p^3 q^2 = \frac{5!}{3!2!}\left(\frac{1}{6}\right)^3\left(\frac{5}{6}\right)^2 = \frac{250}{7\,776} = 0{,}032.$$

يمكننا الحصول على الاحتمالات الأخرى ، مع أقلّ ما يمكن من الحسابات ، باستعمالنا العلاقة التي تجمع بين احتمالين متتاليين (أنظر ص 73) .

$$\frac{P_{x+1}}{P_x} = \frac{n-x}{x+1}\cdot\frac{p}{q}.$$

وهكذا :

$$\frac{P_4}{P_3} = \frac{2}{4}\cdot\frac{1}{5} = \frac{1}{10} \quad \text{إذن} \quad P_4 = \frac{1}{10}P_3$$

$$\frac{P_3}{P_2} = \frac{3}{3}\cdot\frac{1}{5} = \frac{1}{5} \quad \text{إذن ،} \quad P_2 = 5P_3 \quad \text{الخ.}$$

هذه ... موضوع الجدول 1 . ونتحقّق من عدم وجود خطأ في الحساب بإجرائنا مجموع الاحتمالات ... الذي يجب أن يساوي واحداً .

ولتجنّب هذه الحسابات ، تمّ وضع جداول للقانون ذي الحدّين ، كبيرة الحجم وتتوقّف على المتغيّرين الوسيطين n وp . تعطي جداول المكتب National Bureau of Standards [1] احتمالات ووظيفة توزيع القانون ذي الحدّين حيث n أصغر من 50 وp

الجدول 1 : حساب احتمالات القانون ذي الحدّين : n=5 ، p=1/6
(القراءة من اليسار إلى اليمين) :

الاحتمال P_x	$\frac{P_{x+1}}{P_x}$	المتغيّرة ذات الحدّين x
3 125/7 776 = 0,402		0
	1	
3 125/7 776 = 0,402		1
	2/5	
1 250/7 776 = 0,161		2
	1/5	
250/7 776 = 0,032		3
	1/10	
25/7 776 = 0,003		4
	1/25	
1/7 776 = 0,000 1		5
7 776/7 776 = 1	—	

تتغيّر كلّ جزء من المئة : p=0,01 ، p = 0,02 الخ . أمّا جداول روميغ(Romig)[2] فتقدّم نفس البيانات حيث n تكون محصورة بين 50 و100 .

هذه الجداول هي من وضع اختصاصيّين . ولكن لحسن الحظ ، كما سنرى لاحقاً ، منذ أن يتجاوز عدد السحوبات n بعض العشرات ، يمكننا تقريب القانون ذي الحدّين بشكل لائق إمّا من قانون بواسّون (Poisson) إمّا من قانون لابلاس - غوس (Laplace-Gauss) ذوي الجداول سهلة الاستعمال .

7 . تسوية قانون ذي حدّين مع توزيع إحصائي ملحوظ

لنفترض بحوزتنا سلسلة من الحالات الملحوظة المتعلّقة بمتغيّرة إحصائية X نجدها منذ البدء مناسبة لشروط تطبيق القانون ذي الحدّين . من الطبيعي أن ينحرف

---

(1) «Tables of the Binomial Probability Distribution (n = 1, 2, ... 49)».National Bureau of Standards , Washington.

(2) H. Romig, 50- 100 «Binomial Tables». John Wiby, New York; Chapman and Hill, London.

التوزيع الملحوظ دائماً ، قليلاً أو كثيراً ، عند التوزيع ذي الحدّين النظري ، إذ تكون الحالات الملحوظة في الواقع مشوبة بتقلّبات عشوائية : ولا تتطابق الترددات التجريبية مع الاحتمالات الناجمة عن القانون ذي الحدّين إلاّ عند حدود سلسلة غير متناهية من الحالات الملحوظة .

بشكل عام ، لا يمكننا منذ البدء تحديد المتغيّر الوسيطي p للقانون ذي الحدّين المناسب للظاهرة الملحوظة: إذ نجهل مكوّنات الوعاء الذي نأخذ منه العيّنة ؛ وغالباً ما يكون تحديد هذا التكوين هدف البحث الإحصائي نفسه . من المفترض إذن أن نسوّي مع التوزيع الملحوظ القانون ذا الحدّين الأقرب ، وتقوم طريقة التسوية ، من أجل تمثيل الظاهرة ، على اعتماد القانون ذي الحدّين حيث الأمل الرياضي يساوي متوسّط التوزيع الملحوظ .

بالتالي ، بعد أن نحسب متوسّط التوزيع الملحوظ x ، نأخذ للمتغيّر الوسيطي p القيمة :

$$p = \frac{\bar{x}}{n}$$

لأنّ أمل القانون ذي الحدّين الرياضي هو : E {X} = np .

مثلاً . نستخدم إحدى الآلات لصنع قطع ميكانيكية ، وينتج عنها عدد معيّن من القطع المعيبة يجب رفضه . نلاحظ مئة عيّنة (N=100) ، تتكوّن كلّ منها من 40 قطعة (n=40) ، نأخذها بالصدفة من الكمّية المصنوعة . وهكذا نحصل على توزيع العيّنات المائة تبعاً لعدد القطع المعيبة الموجودة في كلّ عيّنة (الجدول 2) .

الجدول 2 . توزيع 100 عيّنة من 40 قطعة تبعاً لعدد القطع المعيبة

عدد القطع المعيبة x_i	عدد العيّنات N_i	$N_i x_i$
0	28	0
1	40	40
2	21	42
3	7	21
4	3	12
5	1	5
6 وأكثر	0	0
المجموع	$\sum N_i = 100$	$\sum N_i x_i = 120$

إذا افترضنا أنّ نسبة القطع المعية p في الكمّية المصنوعة تبقى ثابتة ، فإنّ عدد القطع للرفض في كلّ عيّنة هو متغيّرة عشوائية ذات حدّين بمتغيّرين وسيطيّين n=40 (مقدار العيّنة) وp الذي نجهل قيمته (نسبة القطع المعية في الكمّية المصنوعة) .

لنحسب متوسّط التوزيع الملحوظ x :

$$\bar{x} = \frac{\sum N_i x_i}{N} = \frac{120}{100} = 1,2 .$$

كي نقدّر p ، سنقيم المعادلة بين أمل القانون ذي الحدّين الرياضي وقيمة هذا المتوسّط :

$$E(X) = np = \bar{x}$$

$40p = 1,2$ إذن ∴ $p = 0,03$.

من الطبيعي أن لا تتطابق الترددات الملحوظة تماماً مع احتمالات القانون ذي الحدّين المسوّى $\mathscr{B}(40; 0,03)$ (الجدول 3) .

سوف نتعرّف لاحقاً (الفصل III ، القسم III) إلى طريقة تسمح لنا بالحكم على نوعية هذه التسوية ، أي تحديد ما إذا كان بالإمكان عزو الانحرافات الملحوظة بين الترددات التجريبية والاحتمالات النظرية إلى التقلّبات العشوائية فقط . وهكذا نتحقّق ما إذا كان بوسعنا اعتبار نسبة القطع المعية p في الكمية المصنوعة ثابتة وتساوي 3% .

ملاحظة : في هذا المثل ، لا يجب الخلط بين المتغيّرين الوسيطين n وN: n هي مقدار كلّ من العيّنات ، N هي عدد هذه العيّنات .

الجدول 3 . مقارنة الترددات الملحوظة مع الاحتمالات المسوّاة .

عدد القطع المعية x	الترددات الملحوظة f_x	الاحتمالات المسوّاة P_x
0	0,28	0,295 7
1	0,40	0,365 8
2	0,21	0,220 6
3	0,07	0,086 4
4	0,03	0,024 7
5	0,01	0,005 5
6	0,00	0,001 0
7	0,00	0,000 1
8 وأكثر	0,00	0,000 2
المجموع	1,00	1,000 0

القسم II

القانون فوق الهندسي

1 . تعريف .ـ 2 . المقاييس : A . الأمل الرياضي ؛ B . التباين .ـ 3 . الميل نحو القانون ذي الحدّين .

إنّ القانون ذا الحدّين يناسب سحب عيّنة مع ردّ من مجتمع إحصائي يتضمّن فئتين من الوحدات الاحصائية أو الأفراد ، بعكس القانون فوق الهندسي الذي يناسب سحب عيّنة دون ردّ . وفي الواقع فإنّه يُعمد عادةً إلى هذه الطريقة الأخيرة من أجل أخذ عيّنة ما : فبالنسبة لعيّنتين متساويتي الحجم ، تعطينا طريقة السحب المستنفد تقديرات أدق (أنظر الفصل VI ، ص 247) . إلاّ أنّ خصائص القانون فوق الهندسي واستعماله أقلّ سهولة من خصائص واستعمال القانون ذي الحدّين . لكن ما أن يصبح مقدار المجتمع الإحصائي N كبيراً بالنسبة لمقدار العيّنة n ، فإنّ القانون فوق الهندسي يصبح قريباً جدّاً من القانون ذي الحدّين ويصبح بالإمكان المعادلة بينهما .

1 . تعريف

لنعد إلى مثل الوعاء الذي يحتوي N كرة ضمن فئتين :

- كرات بيضاء B بنسبة p ،
- كرات حمراء R بنسبة $q = 1 - p$

نجري n سحباً متتالياً لإحدى الكرات ، دون ردّها إلى الوعاء قبل السحوبات اللاحقة ، أو ، والنتيجة هي نفسها ، نأخذ دفعة واحدة عيّنةً تتكوّن من n كرة . نحدّد المتغيّرة العشوائية فوق الهندسية X كعدد الكرات البيضاء الحاصلة خلال السحوبات الـ n .

قانون الاحتمال

في حالة القانون ذي الحدّين ، وبسبب ردّ الكرة إلى الوعاء ، كانت السحوبات المتتالية مستقلّة ، الأمر يختلف بالنسبة للمتغيّرة فوق الهندسية : فاحتمال أن نسحب كرة بيضاء عند السحب رقم i يتوقّف على نتيجة السحوبات المتقدّمة . ففي الحقيقة يتغيّر تكوين الوعاء تدريجياً خلال التجارب ، حيث يُستنفد مقدار الوعاء رويداً رويداً ، ومن هنا تسمية السحب المستنفد التي أعطيناها لهذا النمط من اختيار العيّنة .

مثلاً . لنأخذ وعاء يحتوي 10 كرات منها 2 بيضاء B و8 حمراء R . متغيّرات

القانون فوق الهندسي الذي يناسب سحب عيّنة من هذا الوعاء الوسيطية هي :

$N = 10$ ، مقدار المجتمع الاحصائي ؛

$p = 0{,}2$ ، نسبة الكرات البيضاء (تكوين الوعاء) ؛

n ، حجم العيّنة .

يمكننا الحصول على مختلف الحوادث الممكنة تبعاً لصورة شجرة ، كما في حالة القانون ذي الحدّين . إلاّ أنّه يجب الانتباه إلى أنّه ، انطلاقاً من السحب الثالث ، قد تستنفذ جميع الكرات البيضاء : لا يمكن للمتغيّرة فوق الهندسية X أن تأخذ سوى القيمة 0 ، 1 أو 2 .

نحصل ، بالنسبة للسحوبات الثلاثة الأولى ، على قوانين الاحتمال الممثّلة أسفله .

عند السحب الثالث مثلاً ، تأخذ المتغيّرة X القيمة 2 لكلّ من الحوادث النموذجية التالية : $B_1B_2R_3$ ، $B_1R_2B_3$ ، $R_1B_2B_3$ ، حيث الإشارة ترمز إلى رتبة السحب .

احتمال الحدث $B_1B_2R_3$ هو ، بفضل قاعدة الاحتمالات المركّبة :

$$P\{B_1 B_2 R_3\} = P\{B_1\}.P\{B_2/B_1\}.P\{R_3/B_1 B_2\}.$$

بعد حصولنا على كرة بيضاء عند السحب الأوّل ، يبقى في الوعاء 9 كرات منها واحدة بيضاء . بالتالي ، فإن احتمال الحصول على كرة بيضاء عند السحب الثاني ، مع العلم أنّنا قد حصلنا على واحدة عند السحب الأوّل هو :

$$P\{B_2/B_1\} = \frac{1}{9}.$$

الحدث النموذجي	المتغيرة العشوائية X	الاحتمال $P\{X\}$
	السحب الأوّل : $n=1$	
B	1	2/10
R	0	8/10
		المجموع : 1
	السحب الثاني : $n = 2$	
BB	2	$\frac{2}{10}\cdot\frac{1}{9} = \frac{1}{45}$

BR RB	1	$\frac{2}{10}\cdot\frac{8}{9}$ $+\frac{8}{10}\cdot\frac{2}{9} = \frac{16}{45}$
RR	0	$\frac{8}{10}\cdot\frac{7}{9} = \frac{28}{45}$
		1

السحب الثالث : n=3

BBR BRB RBB	2	$\frac{2}{10}\cdot\frac{1}{9}\cdot\frac{8}{8}$ $+\frac{2}{10}\cdot\frac{8}{9}\cdot\frac{1}{8}$ $+\frac{8}{10}\cdot\frac{2}{9}\cdot\frac{1}{8}$ $= \frac{3}{45}$
BRR RBR RRB	1	$\frac{2}{10}\cdot\frac{8}{9}\cdot\frac{7}{8}$ $+\frac{8}{10}\cdot\frac{2}{9}\cdot\frac{7}{8}$ $+\frac{8}{10}\cdot\frac{7}{9}\cdot\frac{2}{8}$ $= \frac{21}{45}$
RRR	0	$\frac{8}{10}\cdot\frac{7}{9}\cdot\frac{6}{8} = \frac{21}{45}$
		المجموع : 1

كذلك :
$$P\{R_3/B_1\,B_2\} = \frac{8}{8} = 1,$$

إذن :
$$P\{B_1\,B_2\,R_3\} = \frac{2}{10}\cdot\frac{1}{9}\cdot\frac{8}{8} = \frac{1}{45}.$$

بنفس الطريقة نحسب :
$$P\{B_1\,R_2\,B_3\} = \frac{2}{10}\cdot\frac{8}{9}\cdot\frac{1}{8} = \frac{1}{45}$$
$$P\{R_1\,B_2\,B_3\} = \frac{8}{10}\cdot\frac{2}{9}\cdot\frac{1}{8} = \frac{1}{45}$$

إذن ، يساوي احتمال أن تأخذ المتغيّرة X القيمة 2 ، وهي القيمة المطابقة لتحقيق حدث أو آخر من هذه الحوادث النموذجية الثلاثة ، 3/45 :
$$P\{X=2\} = P\{B_1\,B_2\,R_3\} + P\{B_1\,R_2\,B_3\} + P\{R_1\,B_2\,B_3\} = 3/45.$$

يمكننا التحقّق ، على الجدول ، أنّ مجموع الاحتمالات يساوي واحداً .

بشكل عام ، عند السحب رقم n ، احتمال أن تأخذ المتغيّرة $\tilde{X}$ القيمة x هو :
$$P_x = P\{X = x\} = \frac{C_{N_p}^{x}\cdot C_{N_q}^{n-x}}{C_N^n}$$

في الحقيقة ، لنعطِ رقماً إلى كلّ من الـ N كرة الموجودة في الوعاء :

$$\underbrace{1, 2, \dots, Np}_{\text{كرات بيضاء}}, \quad \underbrace{Np+1, \dots, N}_{\text{كرات حمراء}}.$$

إذا تمّ سحب العيّنة بالصدفة أي عشوائياً ، فإنّ كلّ التوافيقيات C_N^n التي يمكننا إنجازها باختيارنا n كرة من N موجودة في الوعاء هي متعادلة الاحتمال : إنّها الإمكانيات المحتملة .

لنعدّد بين هذه الأخيرة الإمكانيات المناسبة لوجود x كرة بيضاء وn−x كرة حمراء . هناك C_{Np}^x طريقة اختيار x كرة بيضاء من Np كرة بيضاء موجودة في الوعاء ، ولكلّ من هذه التوافيقيات تطابق C_{Nq}^{n-x} طريقة أخذ الـ n−x كرة حمراء المتمّمة من ضمن Nq كرة حمراء موجودة في الوعاء . هناك إذن ، بالإجمال :

$$C_{Np}^x \cdot C_{Nq}^{n-x}$$

إمكانية مناسبة للحصول على x كرة بيضاء .

لا يمكن لعدد الكرات البيضاء x في العيّنة أن يأخذ قيمة أكبر من مقدار العيّنة n أو من عدد الكرات البيضاء Np الموجودة في الوعاء :

أصغر $x \leqslant (n, Np)$ (x أصغر من أصغر (n, Np)) .

ويصحّ نفس التفكير بالنسبة للـ n−x كرة حمراء في العيّنة :

أصغر $n-x \leqslant (n, Nq)$ (n−x أصغر من أصغر (n, Nq)) ،

إذن : أكبر $x \geqslant (0, n-Nq)$ (x أكبر من أكبر (0, n−Nq)) .

أخيراً :

أصغر $(n, Np) \geqslant x \geqslant$ أكبر $(0, n-Nq)$.

باختصار ، إنّ المتغيّرة فوق الهندسية X هي متغيّرة عشوائية منفصلة تتعلّق بثلاثة متغيّرات وسيطية :

N ، مقدار المجتمع الإحصائي

p ، نسبة الكرات البيضاء البدائية في هذا المجتمع الإحصائي ،

n ، عدد السحوبات المتتالية (مقدار العيّنة) .

قيم هذه المتغيّرة الممكنة هي :

أصغر $(n, Np) \geqslant x \geqslant$ أكبر $(0, n-Nq)$.

واحتمال القيمة x هو :

$$P\{X=x\}=\frac{C_{Np}^{x}\cdot C_{Nq}^{n-x}}{C_N^n}$$

لنشير إلى أنّه إذا كان مقدار العيّنة n في الوقت نفسه أصغر من مقدار الكرات الحمراء Nq ، فإنّ القيم الممكنة هي ، كما في حالة القانون ذي الحدّين : $0 \leq x \leq n$.

2 . مقاييس القانون فوق الهندسي

A . الأمل الرياضي

إنّ أمل التوزيع فوق الهندسي الرياضي (أو متوسّطه أو معدّله الوسطي) يساوي : np

$$E\{X\}=np$$

إذن للقانون ذي الحدّين والقانون فوق الهندسي نفس الأمل الرياضي .

البرهان : انطلاقاً من تعريف الأمل الرياضي :

$$E\{X\}=\sum_x x.P_x=\sum_x x\frac{C_{Np}^{x}\cdot C_{Nq}^{n-x}}{C_N^n}.$$

لنوسّع العبارة :

$$E\{X\}=\sum_x x\frac{n!\,(N-n)!}{N!}\cdot\frac{Np!}{x!\,(Np-x)!}\cdot\frac{Nq!}{(n-x)!\,(Nq-n+x)!}$$

$$=\sum_{x=1}\frac{(N-n)!\,(n-1)!\,n}{(N-1)!\,N}\cdot\frac{(Np-1)!\,Np}{(x-1)!\,(Np-x)!}\cdot\frac{Nq!}{(n-x)!\,(Nq-n+x)!}$$

نضع np كعامل مشترك ونكتشف إذن تحت رمز الجمع Σ عبارات تعداد التوافيقيات :

$$E\{X\}=np\sum_{x=1}\frac{C_{Np-1}^{x-1}\cdot C_{Nq}^{n-x}}{C_{N-1}^{n-1}}.$$

لنجرِ استبدال المتغيّرات التالي :

$$x'=x-1,\qquad N'=N-1,\qquad n'=n-1$$

$$N'p'=Np-1,\qquad N'q'=Nq.$$

فنحصل على:

$$E\{X\}=np\sum_{x'}\frac{C_{N'p'}^{x'}\cdot C_{N'q'}^{n'-x'}}{C_{N'}^{n'}}.$$

لكن

$$\sum_{x'}\frac{C_{N'p'}^{x'}\cdot C_{N'q'}^{n'-x'}}{C_{N'}^{n'}}$$

تمثّل مجموع احتمالات قانون فوق هندسي ذي متغيّرات N′ ، p′ وn′ : هذا

المجموع يساوي واحداً .

B . التباين

إنّ تباين التوزيع فوق الهندسي يساوي $\frac{N-n}{N-1}.npq$:

$$V\{X\} = \frac{N-n}{N-1}.npq .$$

منذ أن نقوم بإجراء أكثر من سحب واحد ، يصبح المعامل $(N-n)/(N-1)$ أصغر من 1 . إذن هذا التباين هو أصغر من تباين القانون ذي الحدّين الذي يساوي npq ، وكلّما يقترب مقدار العيّنة من مقدار المجتمع الإحصائي ، فإنّ تباين المتغيّرة فوق الهندسية يصغر ، وهذا أمر طبيعي . في غاية الأمر، نسحب كلّ المجتمع الإحصائي ويصبح التباين يساوي صفراً : حيث نعرف تماماً عدد الكرات البيضاء الموجودة في الوعاء .

لكن ، عندما يكون مقدار المجتمع الاحصائي N كبيراً بالنسبة لحجم العيّنة n ، فإنّ المعامل موضع الكلام لا يختلف كثيراً عن 1 :

$\frac{N-n}{N-1}$ يميل إلى 1 عندما تميل N إلى ما لا نهاية

عندئذٍ تصبح طريقتا سحب العيّنة ، المستنفدة (القانون فوق الهندسي) ومع ردّ (القانون ذو الحدّين) متعادلتين ، كما سنرى في الفقرة اللاحقة .

لهذه النتيجة أهمّية كبرى في تطبيق الأبحاث الإحصائية عملياً . ففي الحقيقة لا تتعلّق دقّة البحث الإحصائي عملياً ، في الحالة الأكثر تردّداً حيث حجم المجتمع الإحصائي كبير وحجم العيّنة صغير نسبياً ، إلاّ بمقدار العيّنة ، وليس بمقدار المجتمع الإحصائي . فإنّ سحب عيّنة من 1000 وحدة إحصائية من مجتمع إحصائي مقداره 100 000 أو 10 000 000 يعطي نفس الفكرة تقريباً عن تكوين هذا المجتمع . بعبارة أخرى ، تتوقّف الدّقة الحاصلة على مقدار العيّنة n أكثر من نسبة البحث الإحصائي n/N ، كما قد يُخيّل لنا .

بالتالي ، يكون الإستقصاء بواسطة البحث الإحصائي أقلّ كلفة ، نسبياً ، بقدر ما يكون المجتمع الإحصائي كبيراً .

البرهان . إنّ حساب التباين يشبه من حيث مبدئه حساب المتوسّط ، يمكننا بادىء ذي بدء حساب العزم العاملي ذي الدرجة 2: $E\{X(X-1)\}$.

$$E\{X(X-1)\}=\sum_x x(x-1)P_x=\sum_x x(x-1)\frac{C^x_{Np}\cdot C^{n-x}_{Nq}}{C^n_N}$$

$$=\sum_x x(x-1)\frac{n!(N-n)!}{N!}\cdot\frac{Np!}{x!(Np-x)!}\cdot\frac{Nq!}{(n-x)!(Nq-n+x)!}$$

$$=\sum_{x=2}\frac{(N-n)!(n-2)!(n-1)n}{(N-2)!(N-1)N}\cdot\frac{(Np-2)!(Np-1).Np}{(x-2)!(Np-x)!}\cdot\frac{Nq!}{(n-x)!(Nq-n+x)!}$$

نضع $np\frac{(n-1)(Np-1)}{N-1}$ كعامل مشترك ونكتشف عندئذٍ تحت رمز الجمع $\sum$ عبارات تعداد التوافيقيات :

$$E\{X(X-1)\}=np\frac{(n-1)(Np-1)}{N-1}\sum_{x=2}\frac{C^{x-2}_{Np-2}\cdot C^{n-x}_{Nq}}{C^{n-2}_{N-2}}.$$

إلّا أنّ هذا المجموع الأخير يساوي 1 ، لأنّه يمثّل مجموع الاحتمالات المنسوبة إلى متغيّرة هندسية ذات متغيّرات وسيطة $N'=N-2,\ p'=\frac{Np-2}{N-2}$ ، $n'=n-2.$

بالتالي :

$$E\{X(X-1)\}=np\frac{(n-1)(Np-1)}{N-1}$$

بفضل خصائص الأمل الرياضي (أنظر الفصل I ، ص 55) :

$$E\{X(X-1)\}=E\{X^2-X\}=E\{X^2\}-E\{X\}.$$

إذن :

$$E\{X^2\}=E\{X(X-1)\}+E\{X\}$$
$$=np\left[\frac{(n-1)(Np-1)}{N-1}+1\right]=np\frac{Nnp+Nq-n}{N-1}.$$

ولكنّنا نذكر أنّه يمكننا التعبير عن التباين بواسطة $E\{X\}$ و$E(X^2)$ ، وهما العزمان من الدرجة الأولى والثانية (أنظر الفصل I ، ص 63) :

$$V\{X\}=E\{X^2\}-[E\{X\}]^2$$
$$=np\frac{Nnp+Nq-n}{N-1}-(np)^2=\frac{N-n}{N-1}npq.$$

3 . ميل القانون فوق الهندسي نحو القانون ذي الحدّين

عندما يصبح مقدار المجتمع الإحصائي N كبيراً جدّاً ، وn وp يبقيان ثابتين ، فإنّ القانون فوق الهندسي يميل نحو القانون ذي الحدّين .

إنّها النتيجة التي ستسمح لنا عملياً بتطبيق القانون ذي الحدّين ، حيث استعماله أسهل بكثير من استعمال القانون فوق الهندسي ، على الأبحاث الإحصائية وإجراءات

التقدير على العيّنة . في الحقيقة ، يتمّ أخذ معظم العيّنات بواسطة السحب المستفيد ، بشكل لا يمكننا معه تعيين الوحدة الاحصائية مرّتين : يجب إذن على وجه الدّقة تطبيق القانون فوق الهندسي .

في الواقع ، بسبب حجم التجمّع الإحصائي المرتفع عامّة ، يبقى احتمال سحب كرة بيضاء قريباً من p على مرّ السحوبات المتتالية ، رغم عدم ردّ الكرة إلى الوعاء .

لنأخذ مثلاً وعاء يحتوي 100 000 كرة ، منها 40 000 بيضاء ، نسحب منه دون ردّ عيّنة من 1 000 كرة :

$$N = 100\ 000, \quad p = 0{,}4, \quad n = 1\ 000$$

عند السحب الأوّل ، احتمال سحب كرة بيضاء هو :

$$P\{B_1\} = \frac{40\ 000}{100\ 000} = 0.4 = p.$$

عند السحب الثاني ، يصبح هذا الاحتمال :

ـ إذا حصلنا على كرة حمراء عند السحب السابق :

$$P\{B_2/R_1\} = \frac{40\ 000}{99\ 999} = 0.400\ 004 \approx 0.4 \quad (^1)$$

ـ إذا حصلنا على كرة بيضاء عند السحب السابق :

$$P\{B_2/B_1\} = \frac{39\ 999}{99\ 999} = 0{,}399\ 994 \approx 0.4 .$$

عند السحب الأخير ، وإذا أخذنا أقلّ الافتراضات مناسبة ، وهو حيث تم سحب كرة بيضاء على طول السحوبات الـ 999 الأولى ، فإن احتمال الحصول على كرة بيضاء هو :

$$P\{B_{1000}/B_1\ B_2 \ldots B_{999}\} = \frac{39\ 001}{99\ 001} = 0{,}394 \approx 0{,}4 .$$

عند كلّ من السحوبات ، يبقى احتمال الحصول على كرة بيضاء إذن قريباً من النسبة البدائية p للكرات البيضاء الموجودة في الوعاء : عملياً ، نجد أنفسنا ضمن شروط تطبيق القانون ذي الحدّين .

---

(1) 0,4 = 0,400 004 تقرأ 0,400004 لا تختلف كثيراً عن 0,4

بشكل أدقّ :

$$P_x = \frac{C^x_{Np}.C^{n-x}_{Nq}}{C^n_N} = \frac{\frac{Np!}{x!(Np-x)!}.\frac{Nq!}{(n-x)!(Nq-n+x)!}}{\frac{N!}{n!(N-n)!}}$$

إذا اختزلنا :

$$P_x = \frac{n!}{x!(n-x)!}.\frac{[Np(Np-1)\dots(Np-x+1)][Nq(Nq-1)\dots(Nq-n+x+1)]}{N(N-1)\dots(N-n+1)}.$$

إلّا أنّه عندما تميل N نحو اللانهاية

$$\underbrace{Np(Np-1)\dots(Np-x+1)}_{x \text{ عاملاً}} \sim (Np)^x \quad (^1)$$

$$\underbrace{Nq(Nq-1)\dots(Nq-n+x+1)}_{(n-x) \text{ عاملاً}} \sim (Nq)^{n-x}$$

$$\underbrace{N(N-1)\dots(N-n+1)}_{n \text{ عاملاً}} \sim N^n.$$

بالتالي ، إذا وضعنا الكبيرات اللامتناهية المعادلة للبحث عن حدّ العبارة :

$$P_x \to \frac{n!}{x!(n-x)!}.\frac{(Np)^x.(Nq)^{n-x}}{N^n} = \frac{n!}{x!(n-x)!}.p^x q^{n-x}$$

عندما تميل N نحو اللانهاية ($N \to \infty$) .

وفيها نتعرّف على عبارة احتمال المتغيّرة ذات الحدّين .

يعتبر تقريب القانون فوق الهندسي من القانون ذي الحدين صالحاً منذ أن تكون نسبة البحث الاحصائي n/N أصغر من 10% .

القسم III

قانون بواسّون

1 . تعريف . ـ 2 . المقاييس : A . المنوال ؛ B . الأمل الرياضي ؛ C . التباين . ـ 3 . ـ شروط التطبيق : A . تقريب القانون ذي الحدّين ؛ B . سياق بواسّون ؛ C . مجموع متغيّرات بواسّون مستقلّة . ـ 4 . حساب الاحتمالات

(1) $Np(Np-1)\dots(Np-x+1) \sim (Np)^x$ تُقرأ : حاصل الضرب $(Np-x+1)\dots$ $Np(Np-1)$ هو كبير لا متناه يعادل $(Np)^x$.

العملي . جداول قانون بواسّـون . ـ 5 . تسوية قانون بواسّون مع توزيع إحصائي ملحوظ .

إنّ قانون بواسّون (Poisson) يناسب وصف حوادث تكون فرص تحقيقها ضعيـ[illegible] في حالة التوزيع ذي الـ[illegible] . من الضروري أن يبقى احتمال تحقيق الحا[illegible] ثابتاً كي يمكـ[illegible] تطبيق القانون .

[illegible]

[illegible]

[illegible] متغيّرة [illegible] تأخذ القيم الصحيحة :

$$x = 0, 1, 2, \dots ,$$

[illegible] التالي :

$$P_x = P(X = x) = \frac{e^{-m} m^x}{x!} ;$$

حيث m هي متغيّر وسيطي إيجابي ;...e=2,71828 هي قاعدة اللوغاريتمات النيبرية (néperien) . سوف نرى في الفقرة 2 أنّ للمتغيّر الوسيطي m ، الذي تتعلّق به متغيّرة بواسّون كلّياً ، معنى خاصّاً : فهو يساوي في آن واحد متوسّط التوزيع وتباينه .

بوسعنا التحقّق من كون مجموع الاحتمالات يساوي واحداً :

$$\sum_{x=0}^{\infty} P_x = \sum_{x=0}^{\infty} \frac{e^{-m} m^x}{x!} = e^{-m}.e^{m} = 1 .$$

في الواقع ، نضع e^{-m} كعامل مشترك ونتعرّف إلى السلسلة التالية :

$$\sum_{x=0}^{\infty} \frac{m^x}{x!} = 1 + \frac{m}{1!} + \frac{m^2}{2!} + \cdots + \frac{m^x}{x!} + \cdots ,$$

التي تساوي e^m .

ونرمز إلى المتغيّرة X بواسطة : $X = \mathcal{P}(m)$ ، لنشير إلى أنّ المتغيّرة العشوائية X تتبع قانون بواسّون (Poisson) ذا متغيّر وسيطي m .

الشكل

إنّ توزيع بواسّون هو توزيع غير متناظر مع انبساط نحو اليمين، ولكنّه يميل إلى

أن يصبح متناظراً (symétrique) عندما تتزايد m : ويقترب عندها من التوزيع الطبيعي (الشكل 24) .

2 . مقاييس قانون بواسّون

A . المنوال

إنّ منوال قانون بواسّون هو القيمة الصحيحة المحصورة بين m−1 وm .

البرهان . المنوال هو قيمة المتغيّرة العشوائية ذات الاحتمال الأعلى ، إنّه العدد الصحيح x حيث :

$$\frac{P_{x-1}}{P_x} < 1 \quad \text{و} \quad \frac{P_{x+1}}{P_x} < 1 .$$

$$\frac{P_{x-1}}{P_x} = \frac{e^{-m} m^{x-1}}{(x-1)!} \cdot \frac{x!}{e^{-m} m^x} = \frac{x}{m} .$$

$$\frac{P_{x+1}}{P_x} = \frac{e^{-m} m^{x+1}}{(x+1)!} \cdot \frac{x!}{e^{-m} m^x} = \frac{m}{x+1}$$

كي تكون x قيمة المنوال ، يجب أن تحقّق في الوقت نفسه :

$$\frac{x}{m} < 1 \quad \text{و} \quad \frac{m}{x+1} < 1,$$

أي :

$$m - 1 < x < m .$$

إذا كانت m عدداً صحيحاً، يوجد قيمتان للمنوال : m−1 وm (أنظر الشكل 24) .

B . الأمل الرياضي

أمل قانون بواسّون الرياضي (أو معدّله الوسطي أو متوسّطه) يساوي m :

$$E\{X\} = m$$

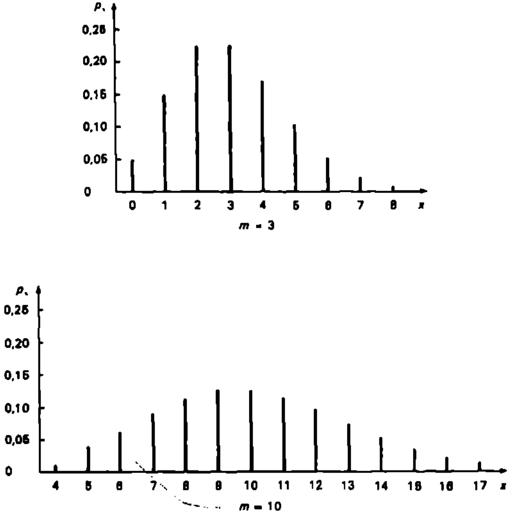
لمتغيّر قانون بواسّون الوسيطي إذن معنى خاص : إنّه متوسّط التوزيع .

البرهان . انطلاقاً من تحديد الأمل الرياضي :

$$E\{X\} = \sum_{x=0}^{\infty} x . P_x = \sum_{x=0}^{\infty} x \frac{e^{-m} m^x}{x!} .$$

بما أنّ أوّل عنصر من المجموع يساوي صفراً يمكننا بدء هذا المجموع عند 1 ووضع m كعامل مشترك :

$$E\{X\} = m \sum_{x=1}^{\infty} \frac{e^{-m} m^{x-1}}{(x-1)!} .$$



الشكل 24 . من أشكال قانون بواسّون

لنجرِ استبدال المتغيّـرة التالي :

$$x' = x - 1 .$$

نحصل على :

$$E(X) = m \sum_{x'=0}^{\infty} \frac{e^{-m} m^{x'}}{x'!}$$

ونكتشف في السلسلة اللامتناهية مجموع احتمالات متغيّرة بواسّون عشوائية يساوي واحداً :

$$\sum_{x'=0}^{\infty} \frac{e^{-m} m^{x'}}{x'!} = 1 .$$

بالتالي : $E(X) = m$

C . التباين

إنّ تباين بواسّون يساوي m : $V(X) = m$

. لمتغيّر قانون بواسّون الوسيطي إذن معنى مزدوج : فهو يساوي في آن واحد متوسّط وتباين التوزيع .

البرهان . حساب التباين هو شبيه من حيث مبدئه بحساب المتوسّط . يمكننا أوّلاً حساب العزم العاملي ذي الدرجة 2 : $E\{X(X-1)\}$

:

$$E\{X(X-1)\} = \sum_{x=0}^{\infty} x(x-1) P_x = \sum_{x=0}^{\infty} x(x-1) \frac{e^{-m} m^x}{x!}$$

العنصران الأوّلان من المجموع يساويان صفراً : إذن يمكننا بدء هذا المجموع عند 2 ووضع m^2 كعامل ضرب مشترك :

$$E\{X(X-1)\} = m^2 \sum_{x=2}^{\infty} \frac{e^{-m} m^{x-2}}{(x-2)!}.$$

لنجرِ استبدال المتغيّرة التالي : $x' = x - 2$. نحصل على :

$$E\{X(X-1)\} = m^2 \sum_{x'=0}^{\infty} \frac{e^{-m} m^{x'}}{x'!}.$$

السلسلة اللامتناهية تساوي 1 لأنّها تمثّل مجموع الاحتمالات المنسوبة إلى متغيّرة بواسّون .

بالتالي :

$$E\{X(X-1)\} = m^2.$$

لكن بفضل خصائص الأمل الرياضي (أنظر الفصل I ، ص 55) :

$$E\{X(X-1)\} = E\{X^2 - X\} = E\{X^2\} - E\{X\},$$

إذن :

$$E\{X^2\} = E\{X(X-1)\} + E\{X\} = m^2 + m.$$

إنطلاقاً من التباين عن التغيّر بواسطة $E\{X\}$ و$E\{X^2\}$ ، وهما العزمان من الدرجة الأولى والثانية (أنظر الفصل I ، ص 63) :

$$V\{X\} = E\{X^2\} - [E\{X\}]^2 = m^2 + m - m^2 = m.$$

3 . شروط التطبيق

يمكن أن نقدّم قانون بواسّون :

- إمّا كحالة خاصّة من القانون ذي الحدّين : فهو القانون الذي يميل نحوه هذا الأخير عندما يصبح عدد التجارب n كبيراً ، بينما يكون احتمال تحقيق

الحدث p ضعيفاً ، لهذا السبب يُدعى قانون بواسّون أحياناً « قانون الأعداد الصغيرة » ؛

- إمّا كنتيجة سياق عشوائي خاص هو سياق بواسّون .

A . تقريب القانون ذي الحدّين بواسطة قانون بواسّون

لنأخذ متغيِّرة عشوائية ذات حدّين $X = \mathscr{B}(n, p)$، حيث المتغيِّر الوسيطي n يكبر بصورة لا متناهية والمتغيّر الوسيطي p يميل نحو صفر بشكل يميل معه حاصل الضرب np نحو ثابتة m . في هذه الشروط ، يميل القانون ذي الحدّين نحو قانون بواسّون بمتغيّر وسيطي m :

$$P_x = C_n^x p^x q^{n-x} \rightarrow \frac{e^{-m} m^x}{x!}.$$

ولهذه النتيجة أهمية الصعيد العملي : فهي تسمح باستبدال القانون ذي الحدّين بقانون بواسّون عندما تكون n كبيرة وp صغيرة وحاصل الضرب np بضع وحدات . التوسيع ذو الحدّين

$$(p + q)^n = \sum_{x=0}^{n} C_n^x p^x q^{n-x}$$

يُستبدَل بالتوزيع اللامتناهي :

$$e^{-m}\left(1 + \frac{m}{1!} + \frac{m^2}{2!} + \cdots + \frac{m^x}{x!} + \cdots\right).$$

بشكل تصبح معه المتغيِّرة X قادرة نظرياً على أخذ عدد غير متناه من القيم الممكنة ، وليس عدداً محدوداً . ففي الحقيقة ، تصبح الاحتمالات وبسرعة صغيرة جدّاً بحيث يمكن تمثيل توزيع متغيِّرة منفصلة متناهية بواسطة قانون بواسّون .

نقبل عادة بوضع قانون بواسّون مكان القانون ذي الحدّين عندما يكون لدينا في آن واحد : $n > 50$ و $p < 10\%$

تكمن أهمّية إمكانية استبدال القانون ذي الحدّين بقانون بواسّون في سهولة استعمال هذا الأخير الكبيرة : فقانون بواسّون لا يتعلّق إلاّ بمتغيّر وسيطي واحد m ، والجداول التي تعطي احتمالات هذا القانون هي جداول بمدخلين (m وx) تملأ بضع صفحات ، بدل الحجم الكبير لجداول القانون ذي الحدّين ذات المداخل الثلاثة : (x،p،n) .

هذا التقارب للقانون ذي الحدّين نحو قانون بواسّون يفسّر وجود هذا الأخير ،

مثلاً ، في الحالات التالية :

- عدد القطع المعيبة في عيّنة كبيرة مأخوذة خلال سياق صناعة بالجملة : بشكل عام ، تكون نسبة القطع المعيبة في مجمل البضاعة ضعيفة .
- عدد الأخطاء المرتكبة خلال جردة عامة لبضاعة تتضمّن عدداً كبيراً من السلع المختلفة ، بشكل عام ، عدد الأخطاء المرتكبة على مرّ سلسلة طويلة من العمليات .

البرهان : لنفترض X متغيّرة عشوائية ذات حدّين :

$$P_x = C_n^x p^x(1-p)^{n-x} = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

لنضع : $np = m + \varepsilon$.

يمكننا الكتابة :

$$P_x = \frac{n(n-1)\ldots(n-x+1)}{x!}\cdot\frac{(np)^x}{n^x}\cdot\frac{1}{(1-p)^x}\cdot\left(1-\frac{np}{n}\right)^n.$$

أي :

$$P_x = \frac{n(n-1)\ldots(n-x+1)}{n^x}\,\frac{(np)^x}{x!}\cdot\frac{1}{(1-p)^x}\left(1-\frac{m+\varepsilon}{n}\right)^n$$
$$= 1\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\frac{1}{(1-p)^x}\frac{(np)^x}{x!}\left(1-\frac{m+\varepsilon}{n}\right)^n$$

عندما $n\to\infty$ و$p\to 0$ ، بشكل يكون معه $np\to m$ حيث m عدد متناه :

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\left(1-\frac{x-1}{n}\right)\to 1$$
$$\frac{1}{(1-p)^x}\to 1$$
$$\frac{(np)^x}{x!}\to\frac{m^x}{x!}$$
$$\left(1-\frac{m+\varepsilon}{n}\right)^n\to e^{-m}.$$

لأنّه كما نعلم $(1-m/n)^n$ تميل نحو e^{-m} عندما تتزايد n بصورة لا متناهية . من ناحية أخرى ، يميل ε نحو صفر .

في هذه الشروط :

$$P_x \to e^{-m}\frac{m^x}{x!}.$$

B . سياق بواسّون

السياق يناسب تحقيق حوادث عشوائية على مرور الزمن ، مثلاً : أعطال في الآلات ، وصول سفن الى مرفأ للتحميل ، اتصالات هاتفية على خطّ معيّن ، وصول زبائن إلى محلّ ما . .

لنفترض أنّ تحقيق حدث خاص (مثلاً ، اتصال هاتفي) يخضع للشروط التالية :

- احتمال تحقيق الحدث خلال فترة قصيرة من الوقت dt هو كمّية متناسبة مع طول هذه الفترة : pdt ؛
- هذا الاحتمال مستقلّ عن عدد الحوادث التي حصلت سابقاً ، ويبقى ثابتاً على طول فترة الملاحظة ؛
- احتمال ظهورين متاليين لهذا الحدث على نفس فسحة الوقت القصيرة dt هو ضئيل جدّاً .

بواسطة هذه الفرضيات ، فإن عدد الحوادث المسجّلة X خلال فسحة من الوقت مدّتها T هو متغيّرة بواسّون عشوائية ذات متغيّر وسيطي $m = pT$.

هذه الخاصّة تفسّر التقاءنا عملياً بقانون بواسّون في كثير من الحالات التي تحقّق الفرضيات السابقة بدرجات متفاوتة من الدّقة . من هذه الحالات :

- وصول سفن إلى مرفأ ، شاحنات إلى مركز تحميل ، طائرات إلى مطار ، زبائن إلى شباك تذاكر ؛
- أعطال الآلات ؛
- الاتصالات الهاتفية ؛
- مبيعات جهاز معيّن في مخزن ، طلب نموذج معيّن لقطعة غيار؛
- بث الذبذبات اللاسلكية ، الخ .

C . مجموع متغيرات بواسّون مستقلّة

مجموع متغيّرتي بواسّون مستقلّتين وبمتغيّرين وسيطيّين m_1 وm_2 ، هو نفسه متغيّرة بواسّون بمتغيّر وسيطي $m = m_1 + m_2$:

$$X_1 = \mathcal{P}(m_1)$$
$$X_2 = \mathcal{P}(m_2)$$
$$Y = X_1 + X_2 = \mathcal{P}(m_1 + m_2) .$$

بالطبع يمكننا بسط هذه النتيجة إلى أي عدد من متغيّرات بواسّون مستقلّة :

$$Z = X_1 + X_2 + \cdots + X_k = \mathcal{P}(m_1 + m_2 + \cdots + m_k).$$

4 . حساب الاحتمالات العملي . جداول قانون بواسّون

يبقى حساب قيمة احتمالات قانون بواسّون العددية ، متعِباً بعض الشيء ، رغم كونه أسهل من حساب القانون ذي الحدّين .

مثلاً . لنأخذ قانون بواسّون ذا المتغيّر الوسيطي $m = 1,2$ ، ولنحسب مثلاً احتمال قيمة المنوال .

المنوال هو القيمة الصحيحة المحصورة بين m−1 وm : إذن يساوي 1 .

$$P_1 = e^{-1,2} \cdot \frac{(1,2)^1}{1!} = 1,2\, e^{-1,2},$$

$$\log P_1 = \log 1,2 - 1,2 \log e = 0,079\ 18 - 1,2 \times 0,434\ 29 = \bar{1},558\ 03\ .$$

(log = لوغاريتم)

إذن :

$$P_1 = 0,361\ 43\ .$$

كما بالنسبة للقانون ذي الحدّين ، يمكننا الحصول على الاحتمالات الأخرى مع أقلّ ما يمكن من الحسابات ، باستعمالنا العلاقة التي تربط بين احتمالين متتاليين :

$$\frac{P_{x+1}}{P_x} = \frac{m}{x+1}$$

هذه الحسابات هي موضوع الجدول 4 ، وهي أسهل بكثير من حسابات العبارة ذات الحدّين المطابقة تماماً (مثلاً n=40 وp=0,03) .

ولكن يوجد جداول تسمح بتجنّب هذه الحسابات . وبما أنّ توزيع بواسّون لا يتعلّق إلاّ بمتغيّر وسيطي واحد m، فإنّ لهذه الجداول مدخلاً مزدوجاً (x,m) واستعمالها أسهل بكثير من جداول القانون ذي الحدّين . ويوجد في ملحق لهذا الكتاب (الجدول 1) جدول حيث m أصغر أو تساوي 15:15, ... ;11 ;10 ;9,5 ;... ;1,5 ;0,0 ,1 ;m= 0,5 أمّا في جداول «Tables for Statisticians and Biometricians» [(1)] ، التي نشرها بيرسون K. Pearson والتي تجمع عدداً كبيراً من المعطيات العددية المفيدة

---

(1) «Tables for Statisticians and Biometricians» ed. by K. Pearson Cambridge Univ. Press.

للحساب الإحصائي ، يوجد جدول لقانون بواسّون حيث m تتغيّر من عشر إلى عشر : $m= 0,1; 0,2; \ldots, 14,9; 15$.

الجداول المعروضة في الملحق تعطينا في آن واحد قيم الاحتمالات P_x ووظيفة التوزيع $F(x)$:

$$P_x = P\{X = x\}, \quad F(x) = P\{X < x\} = P_0 + P_1 + \cdots + P_{x-1}.$$

الجدول 4 . حساب احتمالات قانون بواسّون : $m = 1,2$.

متغيّرة بواسّون x	$\frac{P_{x+1}}{P_x}$	الاحتمال P_x
0		0,301 19
	6/5	
1		0,361 43
	3/5	
2		0,216 86
	2/5	
3		0,086 74
	3/10	
4		0,026 02
	6/25	
5		0,006 25
	1/5	
6		0,001 25
	12/70	
7		0,000 21
	3/20	
8		0,000 03
9 وأكثر		0,000 02
المجموع		1,000 00

هكذا إذا كانت m=6 ، فإنّ احتمال أن تأخذ المتغيّرة العشوائية القيمة 5 هو : $P_5=0,1606$ واحتمال أن تأخذ قيمة أصغر من 5 (5 غير محسوبة) : $F(5)= 0,2851$.

5 . تسوية قانون بواسّون مع توزيع إحصائي ملحوظ

إنّ مبدأ هذه التسوية هو نفسه كما بالنسبة للقانون ذي الحدّين : من أجل تمثيل الظاهرة نعتمد قانون بواسّون يكون أمله الرياضي مساوياً لمتوسّط التوزيع الملحوظ .

مثلاً : لنعد إلى المثل المعروض في موضوع القانون ذي الحدّين (القسم I ، ص

79) : توزيع 100 عيّنة من 40 قطعة مصنوعة بالجملة حسب عدد القطع المعيبة .

يبدو تقريب القانون ذي الحدّين نحو قانون بواسّون ممكناً : إذا كان مقدار كل عيّنة قليلاً بعض الشيء (40 وحدة إحصائية بينما كنا قد قلنا كقاعدة عامّة أن هذا العدد يجب أن يفوق 50 كي يصبح الاستبدال صالحاً) ، فإنّ نسبة القطع المعيبة تبدو صغيرة كفاية كي تكون في النهاية دقّة تقدير الاحتمالات بواسطة قانون بواسّون مناسبة .

الجدول 5. مقارنة التردّدات الملحوظة مع الاحتمالات المسوّاة
(القانون ذو الحدّين وقانون بواسّون) . (القراءة من اليسار إلى اليمين) .

عدد القطع المعيبة x	التردّدات الملحوظة f_x	الاحتمالات P_x: القانون ذو الحدّين	الاحتمالات P_x: قانون بواسّون
0	0,28	0,295 7	0,301 2
1	0,40	0,365 8	0,361 4
2	0,21	0,220 6	0,216 9
3	0,07	0,086 4	0,086 7
4	0,03	0,024 7	0,026 0
5	0,01	0,005 5	0,006 2
6	0,00	0,001 0	0,001 2
7	0,00	0,000 1	0,000 2
8 وأكثر	0,00	0,000 2	0,000 2
المجموع	1,00	1,000 0	1,000 0

متوسّط التوزيع الملحوظ هو :

$$\bar{x} = 1,2$$

احتمالات قانون بواسّون ذي المتغيّر الوسيطي $m = 1,2$ ، المحسوبة في الفقرة السابقة ، هي في الواقع قوية جداً من العبارة الدقيقة لاحتمالات القانون ذي الحدّين حيث $n = 40$ و $p = 0,03$ $(np = 1,2)$ (الجدول 5) .

كما بالنسبة للقانون ذي الحدّين ، يجدر الحكم على نوعية هذه التسوية ببحثنا عمّا إذا كان يمكن بحق إرجاع الانحرافات أو الفروقات الملحوظة بين التردّدات التجريبية والاحتمالات النظرية إلى التقلّبات العشوائية (أنظر الفصل III ، القسم III) .

الفصل الثالث

قوانين التوزيع الإحصائي
النماذج المتواصلة

القسم I

القانون الطبيعي

1 . تعريف : A . قانون الاحتمال الطبيعي ؛ B . قانون الاحتمال الطبيعي المختصر ، C . الشكل .ـ 2 . مقاييس القانون الطبيعي : A . المنوال ؛ B . الأمل الرياضي ؛ C . التباين .ـ 3 . شروط التطبيق : A . نظرية الحدّ المركزي؛ B . تقريب القانون ذي الحدّين ؛ C . قانون متوسّط عيّنة كبيرة ؛ D . مجموع متغيّرات طبيعية مستقلّة . ـ 4 . استعمال جداول القانون الطبيعي : A . جدول كثافة الاحتمال ؛ B . جدول وظيفة التوزيع .ـ 5 . تسوية قانون طبيعي مع توزيع إحصائي ملحوظ : A . التسوية التحليلية ؛ B . التسوية البيانية (الخطّية) : خط هنري . 6. قانون مشتقّ: القانون اللوغطبيعي . A. قانون الاحتمال؛ B . مقاييس القانون اللوغ ـ طبيعي ؛ C . إيجاد الاحتمالات عملياً ؛ D . شروط التطبيق ؛ E . تسوية قانون لوغ ـ طبيعي مع توزيع إحصائي ملحوظ ؛ F . تعميم القانون اللوغ ـ طبيعي .

القانون الطبيعي أو قانون لابلاس ـ غوس (Laplace-Gauss) هو من التوزيعات التي كثيراً ما نلتقي بها عملياً . إنّه ، في الواقع ، القانون الذي يُطبّق على متغيّرة إحصائية تكون نتيجة عدد كبير من الأسباب المستقلّة ، تجتمع تأثيراتها ولا يرجح أحدها على الأخرى . من الواضح أنّها شروط نلتقيها دائماً: أخطاء قياس معينّ ، أقطار قطع

مستديرة مصنوعة بالجملة ، آماد مسار معيّن ، تقلّبات عرضية لكمّية اقتصادية (انتاج ، مبيعات ، الخ .)، الخ . بصورة خاصّة ، يبدو القانون الطبيعي كتقريب للقانون ذي الحدّين عندما يكون مقدار العيّنة كبيراً . تستعمل هذه النتيجة باستمرار على الصعيد العملي ، بشكل خاص في تطبيقات طريقة الأبحاث الإحصائية ، لأنها تسهّل الحسابات بدرجة كبيرة .

1 . تعريف

A . قانون الاحتمال الطبيعي (المعتدل)

المتغيّرة العشوائية الطبيعية X هي متغيّرة متواصلة تأخذ أي قيمة بين ناقص ما لا نهاية (∞−) وزائد ما لا نهاية (∞+) ، وكثافة احتمالها هي :

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right]^{(1)} \quad ;$$

حيث ..$\pi = 3{,}14159$ ؛ $e = 2{,}718\ 28...$ (قاعدة اللوغاريتمات النبيرية) ؛ m و σ هما متغيّران وسيطيان ، m إيجابي أو سلبي و σ إيجابي: سنرى لاحقاً (الفقرة 2) أنّ m يساوي الأمل الرياضي (أو المتوسّط) و σ يساوي الانحراف النموذجي للتوزيع . إذن تحدّد المتغيّرة الطبيعية كلّياً بواسطة متوسّطها m وانحرافها النموذجي σ .

أمّا وظيفة التوزيع ، التي تمثّل احتمال أن تأخذ المتغيّرة العشوائية X قيمة أصغر من x ، فهي :

$$F(x) = P\{X < x\} = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx\,.$$

ونرمز بواسطة : $X = \mathcal{N}(m, \sigma)\,,$

للدّلالة على أن المتغيّرة العشوائية X تتبع قانوناً طبيعياً ذا متغيّرين وسيطيين m و σ .

---

(1) نستعمل العبارة « exp [ ] » عندما يكون قياس الدالّة الأسّية (exponentielle) أطول من مجرّد مجموعة رموز :

$$\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] = e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

بما أنّ القانون الطبيعي يتوقّف على متغيّرين وسيطين ، قد نعتقد أن لجداول هذا القانون ثلاثة مداخل (m ، σ و x) وقد تكون بالتالي ، مثل جداول القانون ذي الحدّين ، كبيرة الحجم ، وغير سهلة الاستعمال . ولكن لحسن الحظ هذا غير صحيح : فمعرفتنا للقانون عند قيمة معيّنة للمتغيّرين m و σ تسمح لنا أن نستنتج بطريقة سهلة توزيعات الاحتمال المناسبة لأيّة قيمة أخرى لـ m و σ .

B . قانون الاحتمال الطبيعي المختصر

لنجرِ استبدال المتغيّرة التالي : $T = \frac{X - m}{\sigma}$

احتمال أن تنتمي X إلى الفسحة اللامتناهية الصغر (x, x+dx) هو :

$$f(x)\,dx = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right] dx .$$

لأنّ : $\frac{x - m}{\sigma} = t, \quad x = \sigma t + m, \quad dx = \sigma\,dt$

بعد استبدال المتغيّرة ، فإنّ احتمال أن تنتمي T إلى الفسحة (t, t+dt) هو :

$$y(t)\,dt = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\,dt .$$

T هي إذن متغيّرة عشوائية طبيعية ، بمتغيّرين وسيطين m=0 و σ =1 . نسمّيها المتغيّرة الطبيعية المركزة المختصرة : مُركزة لأنّ مصدرها (نقطة انطلاقها) هو المتوسّط m ، ومختصرة لأنّنا لقياسها نأخذ الانحراف النموذجي σ كوحدة قياس . ونقول أيضاً المتغيّرة المضبوطة (متوسّطها يساوي صفراً وانحرافها النموذجي واحداً) .

بواسطة استبدال المتغيّرة هذا تُردّ جميع التوزيعات الطبيعية إلى نوع واحد : توزيع المتغيرة الطبيعية المركزة المختصرة .

كثافة احتمال المتغيرة الطبيعية المركزة المختصرة هي :

$$y(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2} , \qquad -\infty < t < +\infty$$

ووظيفة توزيعها التي نشير إليها بواسطة Π(t) هي :

$$\Pi(t) = P\{T < t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2}\,dt .$$

يمكننا التحقّق من أنّ مجموع الاحتمالات يساوي واحداً :

$$\Pi(+\infty) = \int_{-\infty}^{+\infty} f(t)\,dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-t^2/2}\,dt = 1 .$$

بالفعل

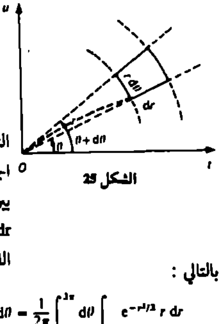
$$[\Pi(+\infty)]^2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-t^2/2}\,dt \times \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-u^2/2}\,du$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \exp\left[-\frac{1}{2}(t^2+u^2)\right] dt\,du .$$

لنجرِ استبدال المتغيّرات التالي (المرور إلى الإحداثيات القطبية) :

$$t = r\cos\theta ,$$
$$u = r\sin\theta .$$

$$t^2 + u^2 = r^2 ,$$
$$dt\,du = r\,dr\,d\theta .$$



الشكل 25

في نظام الإحداثيات الجديد هذا ، النموذج التفاضلي لمساحة المستطيل لا متناهي الصغر ذي الجانبين dt وdu هو ، في الواقع ، المساحة المحصورة بين الدائرة ذات الشعاع r والدائرة ذات الشعاع r+ dr والخط ذي الزاوية القطبية θ والخط ذي الزاوية القطبية $\theta + d\theta$ (الشكل 25) .

بالتالي :

$$[\Pi(+\infty)]^2 = \frac{1}{2\pi}\int_{\theta=0}^{2\pi}\int_{r=0}^{r} e^{-r^2/2}\,r\,dr\,d\theta = \frac{1}{2\pi}\int_0^{2\pi} d\theta \int_0 e^{-r^2/2}\,r\,dr$$

لكن :

$$\int_0^{r} e^{-r^2/2}\,r\,dr = \int_0^{r} e^{-r^2/2}\,d\,\frac{r^2}{2} = [-e^{-r^2/2}]_0^r = 1$$

$$\int_0^{2\pi} d\theta = [\theta]_0^{2\pi} = 2\pi ,$$

إذن :

$$[\Pi(+\infty)]^2 = 1 , \qquad \Pi(+\infty) = 1 .$$

C . الشكل

منحنى كثافة احتمال قانون لابلاس ـ غوس هو منحنى متماثل ذو منوال واحد ، ويتّصل فرعاه الأقصيان تماساً مع محور الإحداثيات السينيات . وقد أعطاه هذا الشكل المميّز اسم منحنى الجرس (الشكل 26) .

وتجتمع الحالات الملحوظة حول المتوسّط على الشكل التالي :

50% ضمن الفسحة $(m - \frac{2}{3}\sigma, m + \frac{2}{3}\sigma)$

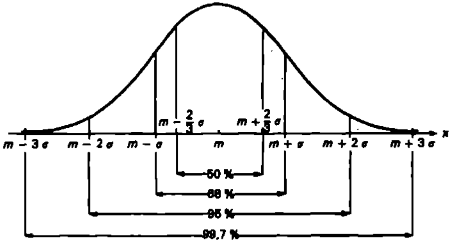
68% ضمن الفسحة $(m - \sigma, m + \sigma)$

95% ضمن الفسحة $(m - 2\sigma, m + 2\sigma)$

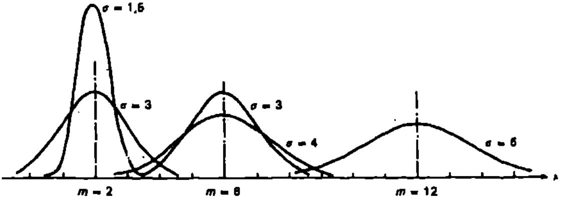
99,7% ضمن الفسحة $(m - 3\sigma, m + 3\sigma)$

عملياً ، تجتمع إذن كل الوحدات تقريباً في فسحة من ستة انحرافات نموذجية حول المتوسّط .

قيمة المتوسّط تحدّد وضعية المنحنى : نستنتج المنحنيات التي لها ذات الانحراف النموذجي من بعضها بواسطة الإزاحة . وحسب قيمة الانحراف النموذجي يكون تشتّت التوزيع (الشكل 27) .



الشكل 26 . شكل القانون الطبيعي :
تجمّع الحالات الملحوظة حول المتوسّط m تبعاً للانحراف النموذجي σ



الشكل 27 . منحنيات كثافة احتمال المتغيّرة الطبيعية
حسب قيم المتغيّرين الوسيطين m و σ

بواسطة استبدال المتغيّرة :

$$T = \frac{X - m}{\sigma},$$

تتحوّل كل هذه المنحنيات إلى المنحنى الذي يمثّل المتغيّرة الطبيعية الممركزة المختصرة (الشكل 28) .

تمثّل وظيفة التوزيع $\Pi(t)$ بواسطة المنجنى التراكمي ،وتطابق نقطة الانعطاف A في هذا المنحنى ، كما في كلّ منحنى تراكمي ، حدّ منحنى كثافة الاحتمال الأقصى ، أي منوال التوزيع . وبما أنّ قيمة وظيفة التوزيع $\Pi(t_0)$ هي مجموع كلّ الاحتمالات النموذجية التي تناسب قيم T الأصغر من t_0 ، فهي تساوي المساحة المخطّطة المحصورة بين منحنى كثافة الاحتمال ومحور الإحداثيات السينيات .

1 . إنّ الدّالة :

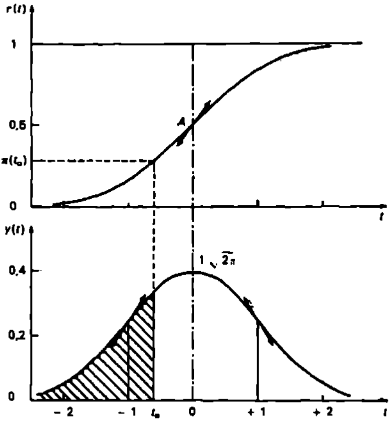
$$\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}.$$

هي دالّة مزدوجة ، أي :

$$\varphi(-t) = \varphi(t).$$

إذن منحنى كثافة الاحتمال هو متناظر بالنسبة للخط ذي الإحداثية السينية t=0 .
وبسبب هذا التناظر :

$$\Pi(-t) = 1 - \Pi(t)$$



الشكل 28 . شكل القانون الطبيعي : منحنى كثافة المتغيّرة الطبيعية المركزة المختصرة ومنحناها التراكمي

والمنحنى التراكمي متماثل بالنسبة لنقطة الانعطاف A (0; 0,5) .

2 . عندما تميل t نحو ∞+ أو ∞− فإنّ y(t) تميل نحو 0 (صفر) :

ـ y(t) هي مقارب (asmyptote) لخط الإحداثيات السينيات عندما $x \to \pm\infty$ ،

ـ Π(t) هي مقارب لخط الإحداثيات السينيات عندما $x \to -\infty$ ،
ومقارب للخط ذي الاحداثية الصادية y=1 عندما $x \to +\infty$

3 . بسبب التماثل فإنّ y(t) تكون حدّاً أقصى عند t = 0 . يمكننا التحقّق أن المشتقّة :

$$y'(t) = -\frac{1}{\sqrt{2\pi}}\, t\, e^{-t^2/2} .$$

تأخذ القيمة صفر عند t = 0 (وكذلك عندما $t \to \pm\infty$)

قيمة الحدّ الأقصى هي : $y(t) = 1/\sqrt{2\pi}$.

وهذا الحدّ الأقصى يُطابق نقطة انعطاف المنحنى $\Pi(t)$.

4 . المشتقّة الثانية لكثافة الاحتمال :

$$y''(t) = \frac{1}{\sqrt{2\pi}}(-e^{-t^2/2} + t^2 e^{-t^2/2}) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}(t^2 - 1),$$

تأخذ القيمة صفر عند $t=\pm1$ ، أمّا المشتقّة الثالثة فهي مختلفة عن الصفر .

لمنحنى كثافة الاحتمال إذن نقطتا انعطاف عند $t=-1$ و $t=+1$.

المتغيّرة الطبيعية ذات المتوسط m والانحراف النموذجي σ والتي نستنتجها من المتغيّرة المركزة المختصرة بواسطة التحوّل الخطّي :

$$x = \sigma t + m,$$

لها منحنى كثافة احتمال متناظر بالنسبة لـ m$(t=0)$ ونقطتا انعطاف عند $x=m-\sigma$ $(t=-1)$ و $x=m+\sigma$ $(t=+1)$.

2 . مقاييس القانون الطبيعي

A . المنوال

المنوال يساوي المتوسّط m بحكم تماثل منحنى الكثافة .

B . الأمل الرياضي

أمل القانون الطبيعي الرياضي (أو متوسّطه) يساوي m :

$$E\{X\} = m$$

لمتغيّر القانون الطبيعي الوسيطي m إذن معنى خاص : فهو متوسّط التوزيع .

البرهان . بحكم التناظر (symétrie) فإنّ أمل المتغيّرة الطبيعية المركزة المختصرة T الرياضي يساوي صفراً .

بالفعل ، انطلاقاً من تعريف الأمل الرياضي :

$$E(T) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t\, e^{-t^2/2}\, dt,$$

وإذا جزّأنا فسحة التكامل ، يمكننا الكتابة :

$$E\{T\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} t\, e^{-t^2/2}\, dt + \frac{1}{\sqrt{2\pi}} \int_{0}^{+\infty} t\, e^{-t^2/2}\, dt.$$

الدّالة (أو الاقتران) $\phi(t) = t\, e^{-t^2/2}$ هي دالّة مفردة ، أي :

$$g(-t) = -g(t).$$

بالتالي :

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} t\,e^{-t^2/2}\,dt = -\frac{1}{\sqrt{2\pi}}\int_{0}^{+\infty} t\,e^{-t^2/2}\,dt.$$

إذن

$$E\{T\} = 0.$$

نستنتج المتغيّرة الطبيعية ذات المتغيّرين الوسيطين m و σ من المتغيّرة المركزة المختصرة بواسطة التحوّل الخطّي :

$$X = \sigma T + m.$$

وبفضل خصائص الأمل الرياضي (أنظر الفصل I ، ص 55) :

$$E\{X\} = \sigma E\{T\} + m,$$

إذن :

$$E\{X\} = m.$$

C . التباين

تباين القانون الطبيعي يساوي σ^2

إذن

$$V\{X\} = \sigma^2.$$

إذن لمتغيّر القانون الطبيعي الوسيطي σ . هو أيضاً ، معنى محدّد جداً : إنّه انحراف التوزيع النموذجي . أخيراً ـ يُحدّد القانون الطبيعي كلّياً عندما نعرف متوسّطه m وانحرافه النموذجي σ .

البرهان .انطلاقاً من تعريف التباين :

$$V\{T\} = E\{(T - E\{T\})^2\} = E\{T^2\}.$$

لأنّ :

$$E\{T\} = 0$$

$$E\{T^2\} = \int_{-\infty}^{+\infty} t^2 f(t)\,dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} t^2 e^{-t^2/2}\,dt.$$

وإذا إعتمدنا التكامل بالتجزئة :

$$\int u\,dv = uv - \int v\,du.$$

$$u = \frac{t}{\sqrt{2\pi}}, \qquad dv = t\,e^{-t^2/2}\,dt,$$

$$du = \frac{dt}{\sqrt{2\pi}}, \qquad v = -e^{-t^2/2}.$$

$$E[T^2] = \frac{1}{\sqrt{2\pi}}[-t\,e^{-t^2/2}]_{-\infty}^{+\infty} + \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-t^2/2}\,dt.$$

العبارة الأولى من المجموع تساوي صفراً والثانية تساوي واحداً ، لأنّها مجموع احتمالات القانون الطبيعي . بالتالي :

$$V(T) = 1$$

انحراف المتغيّرة الطبيعية المركزة المختصرة يساوي واحداً .

نستتج المتغيّرة الطبيعية ذات المتغيّرين الوسيطيّين m وσ من المتغيّرة المركزة المختصرة بواسطة التحوّل الخطّي :

$$X = \sigma T + m.$$

وبفضل خصائص التباين (أنظر الفصل I ، ص 61) :

$$V(X) = \sigma^2 V(T).$$

إذن :

$$V(X) = \sigma^2$$

3 . شروط التطبيق

A . نظرية الحدّ المركزي

إنّ ميدان تطبيق القانون الطبيعي واسع جدّاً ، فهو في الحقيقة حصيلة جمع أسباب تقلب عديدة ومستقلّة ، وذلك تحت شروط غير مقيّدة كثيراً . وتكمن في هذا الأمر نظرية مهمّة جدّاً في حساب الاحتمالات هي نظرية الحدّ المركزي (théorème central-limit) .

لنأخذ متالية المتغيّرات العشوائية $X_1, X_2, \ldots, X_n$ ، التي تناسب عوامل التقلّب المختلفة وتحقّق الشروط التالية :

1 . المتغيّرات X_i هي مستقلّة ؛

2 . آمالها الرياضية $m_1, m_2, \ldots, m_n$ ، وتبايناتها V_1 ، V_2 ، . . . ، V_n ، جميعها موجودة .

3 . نسبة تباين عنصر معيّن من المتالية على مجموع التباينات :

$$\frac{V_i}{\sum_{i=1}^{n} V_i}.$$

يميل نحو الصفر عندما تتزايد n بصورة غير متناهية .

لنسمِّ X مجموع هذه الـ n متغيِّرة عشوائية :

$$X = \sum_{i=1}^{n} X_i .$$

بفضل خصائص الأمل الرياضي (أنظر الفصل I ، ص 55) ، فإنّ أمل X الرياضي يساوي مجموع آمال المتغيِّرات X_1 ، X_2 ، . . . ، X_n ، الرياضية :

$$E(X) = E\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} E(X_i) = m ,$$

حيث :

$$m = m_1 + m_2 + \cdots + m_n .$$

كذلك ، بما أنّ المتغيِّرات X_i مستقلّة ، فإنّ تباين X يساوي مجموع التباينات (خصائص التباين ، الفصل I ، ص 61) :

$$V(X) = V\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} V(X_i) = \sigma^2 ,$$

حيث :

$$\sigma^2 = V_1 + V_2 + \cdots + V_n .$$

يمكننا إذن تأويل الشرط 3 على النحو التالي : إنّ نسبة التغيّر العائدة إلى عامل معيّن للتقلّب هي ضعيفة بالنسبة لنسبة تغيّر X الكلّية ، العائدة إلى مجموع العوامل .

لنشكّل المتغيِّرة المركزة المختصرة :

$$\frac{\sum_{i=1}^{n} X_i - E\left\{\sum_{i=1}^{n} X_i\right\}}{\sqrt{V\left\{\sum_{i=1}^{n} X_i\right\}}} = \frac{X - m}{\sigma} .$$

تؤكّد نظرية الحدّ المركزي أنّ هذه المتغيِّرة تميل إلى أن تتبع القانون الطبيعي المتمركز المختصر عندما تتزايد n بصورة غير متناهية ، مهما تكن قوانين الاحتمال التي تتبعها المتغيِّرات $X_1, X_2, \ldots, X_n$.

نستنتج أنّه يمكننا تمثيل الظواهر التي تُعتبر حصيلة عدد كبير من أسباب تقلّب نموذجية تعمل بصورة مستقلّة ، بواسطة القانون الطبيعي : وهكذا فإن مقاييس (قطر ، وزن ، . . .) قطع تصنع بالجملة تخضع لعدد كبير من أسباب الإخلال : هزّات طفيفة ، تغيّرات في الحرارة ، اختلافات في تجانس المادّة الأوّلية ، الخ . . . ونستنتج

فعلياً على الصعيد العملي أنّ هذه المقاييس غالباً ما تكون موزّعة طبيعياً (حسب القانون الطبيعي) . كذلك الأمر بالنسبة لقياس كمّية معيّنة ، أو مدّة اجتياز مسافة معيّنة أو تقلّبات كمّية اقتصادية معيّنة ، الخ .

إلاّ أنّه لا يجب الاعتقاد أنّ للقانون الطبيعي ميزة شاملة : فقد لا تتوفّر لجميع الشروط المذكورة أعلاه . بشكل خاص ، قد يكون عدد أسباب التقلّب التي تؤثّر على الظاهرة ضعيفاً جدّاً ، أو قد لا تكون تأثيرات هذه الأسباب ممكنة الإضافة بعضها إلى بعض .

وتتحقّق شروط تطبيق القانون الطبيعي في حالتين خاصّتين مهمّتين جدّاً خاصّة في ما يتعلّق بالاستعمال الناتج عنهما في تأويل النتائج الحاصلة عن طريقة الأبحاث الإحصائية :

- تقريب القانون ذي الحدّين من القانون الطبيعي ،
- قانون متوسّط عيّنة كبيرة .

B . تقريب القانون ذي الحدّين من القانون الطبيعي

لنأخذ متغيّرة عشوائية ذات حدّين $X = \mathscr{B}(n, p)$ يتزايد متغيّرها الوسيطي n بصورة غير متناهية ، ولا يكون p قريباً من صفر ولا من 1 . في هذه الشروط ، يميل القانون ذو الحدّين نحو القانون الطبيعي ذي المتغيّرين الوسيطيّين $m = np$ و $\sigma = \sqrt{npq}$

$$\mathscr{B}(n, p) \to \mathscr{N}(np, \sqrt{npq})\,.$$

لهذه النتيجة أهمّية كبيرة على الصعيد العملي : إذا لم يكن p قريباً جدّاً من صفر أو من 1 ، فهو يسمح باستبدال القانون ذي الحدّين بالقانون الطبيعي منذ أن يصبح المتغيّر الوسيطي n مساوياً لبضع عشرات . ومن الطبيعي أن يكون التقريب أفضل كلّما اقترب كلّ من p و q من $\frac{1}{2}$ ، حيث يقترب القانون ذو الحدّين ، الذي يكون عندها متناظراً ، من القانون الطبيعي ، المتناظر هو أيضاً ، بسرعة أكبر .

خلال هذه العملية ، تُستبدل متغيّرة منفصلة تأخذ فقط عدداً محدوداً من القيم بمتغيّرة متواصلة يكون حقل تغيّرها نظرياً غير متناه . في الحقيقة تصبح الاحتمالات وبسرعة صغيرة جداً (يمكن اسقاطها) عندما تميل المتغيّرة الطبيعية نحو $+\infty$ أو $-\infty$ ، بحيث يمكن تمثيل ظاهرة متناهية بواسطة قانون طبيعي . من جهة أخرى ، يستلزم المرور من متغيّرة منفصلة إلى متغيّرة متواصلة بعض الاحتياطات سنذكرها عند عرضنا لاستعمال جداول القانون الطبيعي عملياً (أنظر الفقرة 4 ، ص 122) .

عادة ، نسمح بتقريب القانون ذي الحدّين من القانون الطبيعي عندما يتجاوز كلّ من حاصلي الضرب np وnq من 15 إلى 20 .

إنّ ميل القانون ذي الحدّين نحو القانون الطبيعي هو نتيجة مباشرة لنظرية الحدّ المركزي .

ففي الواقع يمكن اعتبار المتغيّرة ذات الحدّين X ذات المتغيّرين الوسيطيّين n وp ، كمجموع n متغيّرة برنولي X_i مستقلّة (أنظر الفصل II ، القسم I ، ص 72) .

$$X = X_1 + X_2 + \cdots + X_n .$$

بحيث تتوفّر شروط نظرية الحدّ المركزي :

1 . المتغيرات X_i هي مستقلّة ؛

2 . آمالها الرياضية وتبايناتها موجودة :

$$E\{X_i\} = p, V\{X_i\} = pq$$

3 . نسبة تباين متغيّرة برنولي معيّنة على مجموع التباينات :

$$\frac{V(X_i)}{\sum_{i=1}^{n} V(X_i)} = \frac{pq}{npq} = \frac{1}{n}$$

تميل نحو صفر عندما تتزايد n بصورة غير متناهية .

إذن تميل المتغيّرة X ، عندما تتزايد n بصورة غير متناهية ، إلى أن تتبع قانوناً طبيعياً متوسّطه :

$$E(X) = E\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} E(X_i) = np ,$$

وتباينه :

$$V(X) = V\left\{\sum_{i=1}^{n} X_i\right\} = \sum_{i=1}^{n} V(X_i) = npq .$$

ويستند البرهان الدقيق لكيفية ميل القانون ذي الحدّين نحو القانون الطبيعي على تقريب العامليات في قاعدة ستيرلينغ (Stirling) :

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\,(1 + \varepsilon(n)) ,$$

في عبارة احتمالات القانون ذي الحدّين :

$$P_x = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

C . قانون متوسّط عيّنة كبيرة

إنّ قانون احتمال متوسّط ($\bar{x}$) عيّنة كبيرة ذات حجم n ، مسحوبة مع ردّ

من مجتمع احصائي ذي متوسّط m وانحراف نموذجي σ ، هو تقريباً قانون طبيعي ذو متوسّط m وانحراف نموذجي $\sigma/\sqrt{n}$ ، مهما كان قانون توزيع X :

$$\bar{X} \rightarrow N(m, \sigma/\sqrt{n}).$$

وتُعتبر هذه النتيجة بشكل عام صحيحة عندما تتجاوز n تقريباً 30 .

هذا الميل لتوزيع متوسّط عيّنة نحو القانون الطبيعي ، مهما كان قانون توزيع المتغيّرة الإحصائية موضع الدراسة ، هو أيضاً ناتج عن نظرية الحدّ المركزية . حيث تجتمع شروط تطبيق هذه النظرية . متوسّط عيّنة حجمها n هو مجموع n متغيّرة عشوائية :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1}{n} + \frac{x_2}{n} + \cdots + \frac{x_n}{n} ;$$

1 . المتغيّرات x_i مستقلّة لأنّنا أجرينا السحوبات مع ردّ ؛

2 . آمالها الرياضية وتغيّراتها موجودة :

$E(x_i) = m$ (متوسّط المجتمع الإحصائي)

$V(x_i) = \sigma^2$ (تباين المجتمع الإحصائي) ؛

3 . نسبة تباين حالة ملحوظة معيّنة على مجموع التغيّرات :

$$\frac{V(x_i)}{\sum_{i=1}^{n} V(x_i)} = \frac{\sigma^2}{n\sigma^2} = \frac{1}{n}.$$

تميل نحو الصفر عندما تتزايد n بصورة غير متناهية .

إذن عندما يتزايد مقدار العيّنة بصورة غير متناهية ، تميل المتغيّرة العشوائية $\bar{x}$ إلى أن تتبع قانوناً طبيعياً متوسّطه :

$$E(\bar{x}) = E\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = m.$$

وتباينه :

$$V(\bar{x}) = V\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} V(x_i) = \frac{\sigma^2}{n}.$$

عندما يكون سحب العيّنة مستنفداً (لا نردّ الوحدات المسحوبة إلى الوعاء) ، فإنّ ميل توزيع المتوسّط نحو القانون الطبيعي يبقى رغم كون شرط استقلالية الحالات الملحوظة لم يعد محترماً تماماً ، ولكن نبرهن (انظر الفصل VI ، ص 244) أنّ انحراف متوسّط العيّنة النموذجي يكون عندها :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

ونشير إلى أنّ المعامل التصحيحي المضروب بانحراف متوسّط العيّنة النموذجي في حالة السحوبات المستقلّة $(\sigma_{\bar{x}} = \sigma/\sqrt{n})$ للحصول على انحراف متوسّط عيّنة مستنفِدة نموذجي $((\sigma/\sqrt{n}\sqrt{(N-n)/(N-1)}))$ هو نفسه الذي يسمح لنا بالمرور من انحراف المتغيّرة ذات الحدّين النموذجي $(\sqrt{npq})$؛ (إلى انحراف المتغيّرة فوق الهندسية النموذجي $(\sqrt{npq}\sqrt{N-n)/(N-1)})$ (أنظر ص 86) .

ولا عجب في ذلك : إذ يمكن اعتبار تردّد متغيّرة ذات حدّين في عيّنة حجمها n كمتوسّط n متغيّرة برنولي مستقلّة ، وتردّد متغيّرة فوق هندسية كمتوسّط n متغيّرة برنولي غير مستقلّة :

$$f_X = \frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i \, .$$

حيث X_i تساوي 1 أو صفراً حسب طبيعة الوحدة الإحصائية المسحوبة .

عندما تتزايد n ، يميل المعامل التصحيحي $(N-n/N-1)$ نحو الصفر . كما هو طبيعي ، فإنّ دقّة تقدير المتوسّط ، كما دقّة تقدير المقدار أو التردّد ، تتزايد تدريجياً كلّما اقترب مقدار العيّنة من مقدار المجتمع الإحصائي .

إلاّ أنّه بشكل عام ، يكون مقدار المجتمع الإحصائي N كبيراً بالنسبة لحجم العيّنة n : عندها لا يختلف المعامل كثيراً عن 1 :

$$N \to \infty \qquad \text{عندما} \qquad \frac{N-n}{N-1} \to 1$$

D . مجموع متغيّرات طبيعية مستقلّة

إنّ مجموع متغيّرتين طبيعيتين مستقلّتين لهما على التوالي المتغيّرات الوسيطية (m_1, σ_1) و (m_2, σ_2) هو نفسه متغيّرة طبيعيّة متوسّطها :

$$m = m_1 + m_2$$

وتباينها :

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \, .$$

يمكننا بالطبع بسط هذه النتيجة إلى أي عدد من المتغيّرات الطبيعية المستقلّة :

$$X_1 = \mathcal{N}(m_1, \sigma_1)$$
$$X_2 = \mathcal{N}(m_2, \sigma_2)$$
$$\cdots\cdots\cdots\cdots\cdots$$
$$X_k = \mathcal{N}(m_k, \sigma_k)$$

$$Z = X_1 + X_2 + \cdots + X_k = \mathcal{N}(m_1 + m_2 + \cdots + m_k ; \sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_k^2}).$$

4 . إيجاد الاحتمالات عملياً : استعمال جداول القانون الطبيعي

بواسطة استبدال المتغيّـرة التالي :

$$T = \frac{X - m}{\sigma},$$

يمكننا تحويل جميع التوزيعات الطبيعية إلى نوع واحد وهو توزيع المتغيّـرة الطبيعية المـركزة المختصـرة T . ومن أجل هـذه المتغيّرة تمّ حساب وظـائف كثافـة الاحتمال والتوزيع ووضعها في جداول لجدونها في ملحق هذا الكتاب .

A . جدول كثافة الاحتمال y(t)

أوّلاً : **الوصف** يعطينا هذا الجدول كثافة الاحتمال y(t) التي تناسب قيماً إيجابية للمتغيّـرة الطبيعية المركزة المختصرة تتغيّـر من عِشر إلى عشر : t=0,0; 0,1; ... 3,9 . نقرأ الآحاد على الأسطر والأعشار على العواميد (الملحق : الجدول 2) .

مثلاً . إذا كانت t = 1,3 فإنّ كثافة الاحتمال هي : y(t) = 0,1714 .

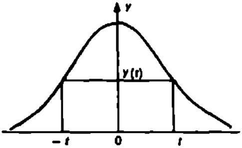
إذا كانت قيم t سلبية

بحكم تماثل لمنحنى الكثافة فإنّ الجدول يسمح بتحديد الكثافات التي تناسب قيماً سلبية لِـ t :

$y(-t) = y(t)$

مثلاً . إذا كانت t=−2,8 ، فإنّ كثافة الاحتمال هي : $y(-2,8) = y(2,8) = 0,0079$

بما أنّ القانون الطبيعي هو توزيع متواصل ، يمكننا الحصول على الكثافات التي تناسب



قيماً لِـ t وسيطة بين القيم الموجودة في الجدول بواسطة الاستكمال الخطّي :

مثلاً . إذا كانت t = 1,36 يمكننا تقدير كثافة الاحتمال على النحو التالي :

$y(1,3) = 0,1714;$

$y(1,4) = 0,1497;$

$$y(1,36) = 0,171\,4 - \frac{(0,171\,4 - 0,149\,7) \times 6}{10} = 0,158\,4 .$$

ولكنّنا نجد جدولاً أدقّ لقيم t تتغيّر كل جزء من المئة : ...; $t = 0,00; 0,01;$ 6,00 ، في الجداول «Tables for Statisticians and Biometricians» التي نشرها بيرسون (K.Pearson) [1] .

ثانياً . الاستعمال : إنّ استبدال المتغيّرة : $T = \frac{X-m}{\sigma}$ يسمح لنا ، بمساعدة الجدول ، بتحديد كثافة الاحتمال التي تناسب أي قيمة للمتغيّرة الطبيعية X ذات المتوسّط m والانحراف النموذجي σ .

بالفعل إذا كانت $X = x$ ، فإنّ كثافة الاحتمال هي :

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right],$$

بينما قيمتها بالنسبة لمتغيّرة طبيعية مركزة مختصرة هي :

$$y(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}.$$

إذن يوجد بين كثافتي احتمال x وt العلاقة التالية :

$$f(x) = \frac{y(t)}{\sigma}$$

مثلاً . لنفترض X متغيّرة طبيعية ذات متوسّط m=5 وانحراف نموذجي $\sigma = 2$. لنبحث عن كثافة الاحتمال التي تناسب $X = 8$ و$X = 4,52$.

إذا كانت $X= 8$ ، فإنّ قيمة المتغيّرة الطبيعية المركزة المختصرة هي :

$$t = (8-5)/2 = 1,5$$

وإذا رجعنا إلى الجدول :

$$y(1,5) = 0,129\,5$$

إذن :

$$f(x) = \frac{y(t)}{\sigma}$$

$$f(8) = \frac{0,129\,5}{2} = 0,064\,8.$$

إذا كانت $X = 4,52$

$$t = (4,52-5)/2 = -0,24,$$

$$y(-0,24) = y(0,24)$$

وبواسطة الاستكمال الخطّي :

$$y(0,20) = 0,391\,0$$

$$y(0,30) = 0,381\,4$$

(1) «Tables for Statisticians and Biometricians», ed. by K. Pearson Cambridge Univ. Press.

$$y(0,24) = 0,391\,0 - \frac{(0,391\,0 - 0,381\,4) \times 4}{10} = 0,387\,2\,.$$

إذن :

$$f(4,52) = \frac{0,387\,2}{2} = 0,193\,6\,.$$

ثالثاً . تقريب القانون ذي الحدّين : يُستعمل الجدول y(t) خاصّة لتقريب احتمالات القانون ذي الحدّين من القانون الطبيعي .

مثلاً . نسحب عيّنة يبلغ حجمها n=40 من مجتمع احصائي يتضمّن النسبة p = 0,4 من الوحدات الاحصائية التي تمثّل الخاصّة A (مثلاً : امتلاك سيّارة) . لنحدّد احتمال أن نلاحظ في العيّنة 20 وحدة تماماً تملك هذه الخاصّة .

عدد الوحدات X التي تمثّل الخاصّة A في العيّنة هو متغيّرة عشوائية ذات حدّين بمتغيّرين وسيطين n = 40 وp = 0,4 :

$$X = \mathcal{B}(40\,;0,4)\,.$$

أمل X الرياضي هو :

$$E\{X\} = np = 16$$

وانحرافها النموذجي :

$$\sigma_x = \sqrt{npq} = 3,1\,.$$

بما أنّ حجم العيّنة n هو كبير بما فيه الكفاية والنسبة p غير قريبة من صفر ولا من 1 ، يمكننا تقريب هذا القانون ذي الحدّين من القانون الطبيعي الذي له نفس الأمل الرياضي والانحراف النموذجي :

$$\mathcal{B}(40\,;0,4) \rightarrow \mathcal{N}(16\,;3,1)\,.$$

المتغيّرة الطبيعية المركزة المختصرة التي تناسب X = 20 هي :

$$t = \frac{20 - 16}{3,1} = 1,29\,.$$

ونجد باعتمادنا جدول القانون الطبيعي والاستكمال الخطّي :

$$y(t) = 0{,}173\,7\ ,$$

إذن :

$$f(x) = \frac{y(t)}{\sqrt{npq}}\ , \qquad f(20) = \frac{0{,}173\,7}{3{,}1} = 0{,}056\,0\ .$$

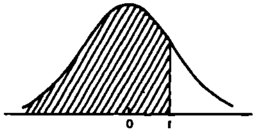
أمّا الاحتمال الصحيح فهو :

$$P\{X = 20\} = C_{40}^{20}\, p^{20}\, q^{20} = \frac{40\,!}{20\,!\,20\,!}(0{,}4)^{20}\,(0{,}6)^{20} = 0{,}055\,4\ .$$

B . جدول وظيفة التوزيع $\Pi(t)$

أولاً . الوصف : يعطينا هذا الجدول لكلّ قيمة إيجابية t للمتغيّرة الطبيعية المركزة المختصرة ، قيمة وظيفة التوزيع المناسبة ، الممثلة على المساحة المخطّطة والتي تساوي احتمال أن تكون T أصغر من t :

$$\Pi(t) = P\{T < t\}\ .$$



وتتغيّر قيم t كلّ جزء من المئة :

$t = 0{,}00; 0{,}01; 0{,}02; \dots$ ، نقرأ الآحاد والأعشار على الأسطر وأجزاء المئة على الأعمدة (الملحق : الجدول 3) .

مثلاً . احتمال أن تكون T أصغر من 1,32 هو :

$$P\{T < 1{,}32\} = \Pi(1{,}32) = 0{,}906\,6\ .$$

احتمال أن تكون T أكبر من t

تمثّل المساحة المحصورة بين المنحنى ومحور الإحداثيات السينية مجموع احتمالات القانون الطبيعي وتساوي واحداً . إذن :

$$P\{T \geqslant t\} = 1 - P\{T < t\} = 1 - \Pi(t)\ .$$

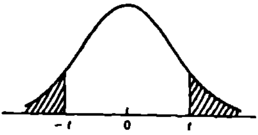
مثلاً . احتمال أن تكون T أكبر من أو تساوي 0,28 هو :

$$P\{T \geqslant 0{,}28\} = 1 - \Pi(0{,}28) = 1 - 0{,}610\,3 = 0{,}389\,7\ .$$

قيم t سلبية

بحكم تماثل المنحنى ، يسمح الجدول بتحديد وظيفة التوزيع لقيم t سلبية :

$$P\{T < -t\} = P\{T \geqslant t\}$$
$$= 1 - P\{T < t\},$$
$$\Pi(-t) = 1 - \Pi(t).$$



مثلاً . احتمال أن تكون T أصغر من 0,77 ـ هو :

$$P\{T < -0{,}77\} = \Pi(-0{,}77) = 1 - \Pi(0{,}77) = 1 - 0{,}779\,4 = 0{,}220\,6.$$

قد يكون لبعض الاستعمالات من الأسهل اعتماد جدول مشتقّ : الجدول P(t) .

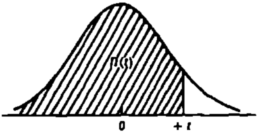
الجدول P(t)

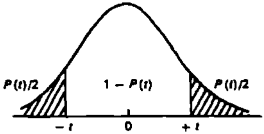
يعطينا الجدول P(t) قيم t حيث يوجد احتمال P أن تكون T خارج الفسحة (−t, +t) . وتتغيّر قيم P كلّ جزء من مئة :

P = 0,00; 0,01; 0,02... (الملحق : الجدول 4) .

يوجد بين π(t) وP(t) العلاقة التالية :

$$P(t) = 2[1 - \Pi(t)].$$





يعطينا الجدول π (t) مباشرة الاحتمالات المناسبة لقيم t معيّنة ، وبالعكس يسمح لنا الجدول P(t) بتحديد سهل لقيم t تناسب قيماً معيّنة للاحتمالات .

مثل 1 . حدّد الفسحة (−t, +t) حيث يساوي احتمال أن تكون T ضمنه 95% :

$$P\{-t \leqslant T < +t\} = 1 - P(t) = 0{,}95$$

$P(t) = 0{,}05$ إذن . $t = 1{,}960\,0$

مثل 2 . حدّد القيمة t حيث يساوي احتمال أن تكون T أصغر منها 90% :

$$P\{T < t\} = \Pi(t) = 1 - \frac{P(t)}{2} = 0{,}90$$

$P(t) = 0{,}20$ إذن $t = 1{,}281\,6$

ثانياً . الاستعمال : يسمح لنا استبدال المتغيّرة التالي :

$$T = \frac{X - m}{\sigma}$$

وبواسطة الجدول بتحديد احتمال أن تكون المتغيّرة الطبيعية X ذات المتوسّط m والانحراف النموذجي σ أصغر من قيمة معطية x ، أو أكبر منها ، أو محصورة بين قيمتين معيّنتين x_1 وx_2

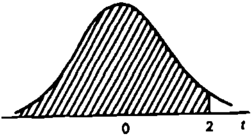
في الواقع :

$$P\{X < x\} = P\{T < t\} = \Pi(t).$$

في كلّ حالة ، يسهّل المخطّط البياني نمط تفكيرنا .

مثلاً . لنفترض X متغيّرة طبيعية بمتوسّط m = 5 وانحراف نموذجي σ=2 .

ـ احتمال أن تكون X أصغر من 9 .



إنّ قيمة المتغيّرة الطبيعية المركزة المختصرة المناسبة لـ X = 9 هي :

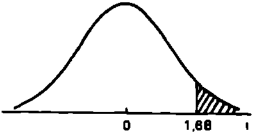
$$t = \frac{9-5}{2} = 2.$$

$$P\{X < 9\} = P\{T < 2\}$$
$$= 0{,}977\,2.$$

ـ احتمال أن تكون X أكبر من أو تساوي 8,36 .

$$t = \frac{8{,}36-5}{2} = 1{,}68$$

$$P\{X \geqslant 8{,}36\} = P\{T \geqslant 1{,}68\}$$
$$= 1 - P\{T < 1{,}68\}$$
$$= 1 - \Pi(1{,}68)$$
$$= 1 - 0{,}953\,5 = 0{,}046\,5\,.$$

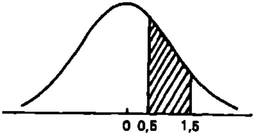


ـ احتمال أن تكون X محصورة بين 6 و8 .

$$t_1 = \frac{6-5}{2} = 0{,}5\,,\quad t_2 = \frac{8-5}{2} = 1{,}5\,,$$
$$P\{6 \leqslant X < 8\}$$
$$= P\{0{,}5 \leqslant T < 1{,}5\}$$
$$= P\{T < 1{,}5\} - P\{T < 0{,}5\}$$

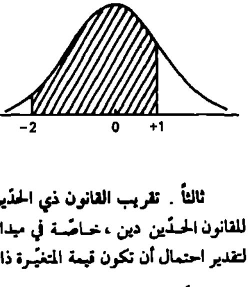
(أنظر الفصل الأوّل ، ص 44) .

$$= \Pi(1{,}5) - \Pi(0{,}5)$$
$$= 0{,}933\,2 - 0{,}691\,5$$
$$= 0{,}241\,7\,.$$



ـ احتمال أن تكون X محصورة بين 1 و7 .

$$t_1 = \frac{1-5}{2} = -2\,,\quad t_2 = \frac{7-5}{2} = 1$$
$$P\{1 \leqslant X < 7\}$$
$$= P\{-2 \leqslant T < 1\}$$
$$= P\{T < 1\} - P\{T < -2\}$$
$$= P\{T < 1\} - [1 - P\{T < 2\}]$$
$$= \Pi(1) - [1 - \Pi(2)]$$
$$= 0{,}841\,3 - 0{,}022\,8 = 0{,}818\,5\,.$$



ثالثاً . تقريب القانون ذي الحدّين . غالباً ما يُعتمد القانون الطبيعي كتقريب للقانون الحـدّين دين ، خاصّة في ميدان الأبحاث الإحصائية ، نستخدم الجدول π(t) لتقدير احتمال أن تكون قيمة المتغيّرة ذات الحدّين داخل فسحة معيّنة .

مثلاً . نسحب عيّنة حجمها n=40 من مجتمع إحصائي يتضمّن النسبة p=0,4 من الوحدات الإحصائية التي تمثّل ميزة معيّنة A . لنقدّر احتمال أن يكون لدينا في العيّنة عدد من الوحدات الإحصائية X التي تمثّل هذه الميزة ، أكبر أو يساوي 16 وقطعاً

أصغر من 20 :

$$P\{16 \leqslant X < 20\}.$$

إنّ عدد الوحدات الإحصائية X التي تمثّل الميزة A هو متغيّرة ذات حدّين أملها الرياضي np = 16 وانحرافها النموذجي $\sqrt{npq}=3{,}1$. يمكننا تقريب هذا القانون ذي الحدّين من قانون طبيعي له نفس الأمل الرياضي ونفس الانحراف النموذجي (راجع المثل ، ص 118) .

بما أنّ الأمر يتعلّق بتقريب متغيّرة منفصلة لا تأخذ سوى قيم صحيحة ، من متغيرة متواصلة ، يجب أن نوجّه عناية خاصّة إلى حدود الفسحة التي نبحث عن احتمالها .

في الواقع ، إذا كانت X متغيّرة متواصلة لا يهمّ كثيراً أن يكون حدّ الفسحة $X<20$ أو $X \leqslant 20$ ، كون احتمال أن تكون X مساوية لـ 20 على وجه الدّقة يساوي صفراً (انظر الفصل I ، ص 45 : فقط احتمال أن تكون X محصورة ضمن فسحة لا متناهية الصغر تحيط بالنقطة ذات الإحداثي السيني 20 له قيمة صغيرة جدّاً ولكن لا تساوي صفراً) .

بالمقابل ، إذا كانت X متغيّرة منفصلة ، فالكتابة : $X<20$ تعني : $X\leqslant 19$ ، كون المتغيّرة X لا يمكنها أخذ أي قيمة بين 19 و20 .

خلال تقريبنا من القانون الطبيعي ، يجب إذن أن نحدّد في الحقيقة :

$$P\{16 \leqslant X \leqslant 19\}.$$

قيمتا المتغيّرة الطبيعية المركزة المختصرة اللتان تناسبان 16 و19 هما :

$$t_1 = \frac{16-16}{3{,}1} = 0, \qquad t_2 = \frac{19-16}{3{,}1} = 0{,}97$$

$$\begin{aligned} P\{16 \leqslant X \leqslant 19\} &= P\{0 \leqslant T \leqslant 0{,}97\} \\ &= P\{T \leqslant 0{,}97\} - P\{T < 0\} \\ &= \Pi(0{,}97) - \Pi(0) \\ &= 0{,}834\,0 - 0{,}500\,0 = 0{,}334\,0 \end{aligned}$$

الاحتمال الحقيقي هو :

$$P\{16 \leqslant X \leqslant 19\} = P_{16} + P_{17} + P_{18} + P_{19},$$

حيث نحسب P_x تبعاً لقاعدة القانون ذي الحدّين :

$$P_x = C_n^x p^x q^{n-x}.$$

فنحصل على :

$$\begin{aligned} P_{16} &= 0{,}127\,9 \\ P_{17} &= 0{,}120\,4 \\ P_{18} &= 0{,}102\,6 \\ P_{19} &= 0{,}079\,2 \\ \hline P\{16 \leqslant X \leqslant 19\} &= 0{,}430\,1. \end{aligned}$$

في هذه الحالة الخاصّة ، لا يبدو التقريب مرضياً بشكل خاص : كما سنرى في ما يلي ، يعود هذا الأمر بدرجة كبيرة إلى أنّنا أهملنا بعض مظاهر تقريب متغيّرة منفصلة من متغيّرة متواصلة .

تصحيح التواصل

بيانياً ، استبدال متغيّرة منفصلة بمتغيّرة متواصلة يعني أن نستبدل مخطّط العيدان بالمدرج التكراري (histogramme) . في هذا المدرج تمثّل الاحتمال P_x بواسطة مستطيل تكون قاعدته ، التي يبلغ طولها واحداً ، مركزة على القيمة x ، أمّا ارتفاعه فيساوي P_x (أنظر الشكل 29) .

بالتالي ، خلال هذا التمثيل ، تمثل مجموع الاحتمالات التالي :

$$P\{16 \leqslant X \leqslant 19\} = P_{16} + P_{17} + P_{18} + P_{19}$$

بواسطة المساحة المخطّطة على الرسم البياني ، أي بواسطة مجموع مساحات المستطيلات المثّلة بين $19 + \frac{1}{2}$ و $16 - \frac{1}{2}$

بشكل عام ، تمثّل الاحتمال :

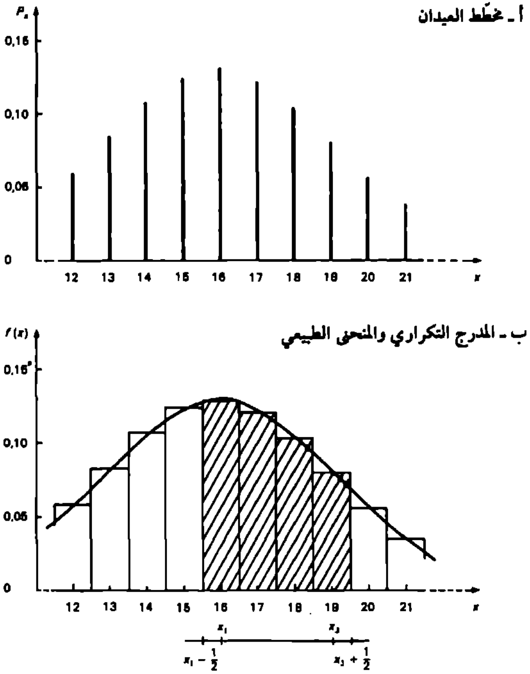
$$P\{x_1 \leqslant X \leqslant x_2\}$$

بواسطة مجموع مساحات المستطيلات المثّلة بين $x_2 + \frac{1}{2}$. و $x_1 - \frac{1}{2}$

خلال تقريبنا من القانون الطبيعي ، نعتمد المنحنى الطبيعي بدلاً من المدرج

التكراري . في الواقـع إذا تحققت جميع شـروط التقارب ، هنـاك تعويض طفيف بـين الأجزاء المضافة إلى أو المحذوفة من كل من المستطيلات . لا يبقى سوى أن يكـون مجموع الاحتمالات محسوباً عـلى الفسحة $(x_1 - \frac{1}{2}, x_2 + \frac{1}{2})$ وليس عـلى الفسحة (x_1, x_2) .

$$P\{x_1 \leqslant X \leqslant x_2\} = F(x_2 + \tfrac{1}{2}) - F(x_1 - \tfrac{1}{2}) .$$



الشكل 29 . تقريب القانون ذي الحدّين من القانون الطبيعي . تصحيح التواصل

يدعى هذا التصحيح لحدود فسحة التكامل تصحيح التواصل . ويأخذ اعتماده أهمّية أكبر كلّما اقترب الحدّان x_1 وx_2 أكثر من المتوسّط m = np وكان الانحراف النموذجي صغيراً أكثر .

مثلاً . لنطبّق تصحيح التواصل على المثال السابق :

$$t_1 = \frac{15{,}5 - 16}{3{,}1} = -0{,}16 , \qquad t_2 = \frac{19{,}5 - 16}{3{,}1} = 1{,}13$$

$$\begin{aligned} P\{16 \leqslant X \leqslant 19\} &= P\{-0{,}16 \leqslant T \leqslant 1{,}13\} \\ &= P\{T \leqslant 1{,}13\} - P\{T < -0{,}16\} \\ &= \Pi(1{,}13) - [1 - \Pi(0{,}16)] \\ &= 0{,}870\,8 - 0{,}436\,4 = 0{,}434\,4 \end{aligned}$$

وإذا قارنّا هذه النتيجة بالاحتمال الحقيقي المحسوب سابقاً (0,4301) ، يبدو لنا التقريب ، هذه المرّة ، مقبولاً لمعظم التطبيقات .

5 . تسوية قانون طبيعي مع توزيع إحصائي ملحوظ

A . التسوية التحليلية

مبدأ هذه التسوية يشبه المبدأ الذي استعملناه من أجل القانون ذي الحدّين وقانون بواسّون : نعتمد لتمثيل الظاهرة القانون الطبيعي (قانون لابلاس - غوس) الذي يكون أمله الرياضي وانحرافه النموذجي مساويين على التوالي لمتوسّط التوزيع الملحوظ وانحرافه النموذجي .

مثلاً : لنأخذ كمّية من 400 برغي (لولب) تتوزّع وحداتها تبعاً لأقطارها حسب معطيات الجدول 6 .

يوحي لنا شكل المدرج التكراري (الشكل 30) بفكرة التسوية مع قانون طبيعي . أجرينا حساب المتوسّط والانحراف النموذجي في الجدول 7 ، حسب الطريقة المعروضة في المجلّد الأول ، الفصل السادس . فحصلنا على :

$$\bar{x} = 3{,}32 \text{ mm}, \qquad \sigma_x = 0{,}10 \text{ mm}$$

إذن نسوّي مع التوزيع قانوناً طبيعياً متغيّراه الوسيطيّان m=3,32 و σ=0,10

$$\mathcal{N}(3{,}32 ; 0{,}10) .$$

عدد البراغي	فئات الأقطار (mm)
3	3,00 إلى أقل من 3,05
6	3,05 إلى أقل من 3,10
13	3,10 إلى أقل من 3,15
23	3,15 إلى أقل من 3,20
39	3,20 إلى أقل من 3,25
78	3,25 إلى أقل من 3,30
91	3,30 إلى أقل من 3,35
72	3,35 إلى أقل من 3,40
42	3,40 إلى أقل من 3,45
17	3,45 إلى أقل من 3,50
9	3,50 إلى أقل من 3,55
5	3,55 إلى أقل من 3,60
2	3,60 إلى أقل من 3,65
400	المجموع

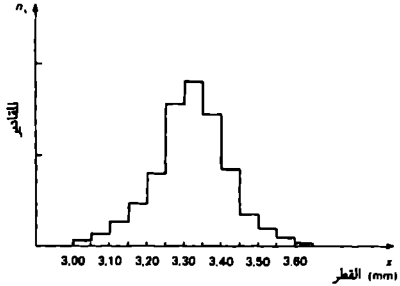
الجدول 6 . توزيع كمّية من 400 برغي حسب أقطارها

حساب المقادير المسوّاة معروض في الجدول 8 . قيم المتغيّرة الطبيعية المركزة المختصرة t_i التي تطابق أطراف الطبقات x_i مذكورة في العمود (2) :

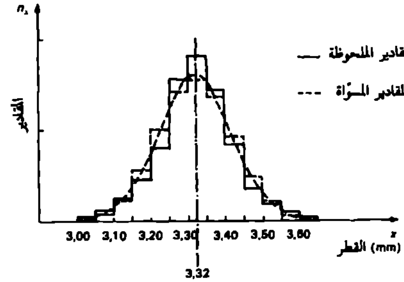
$$t_i = \frac{x_i - m}{\sigma} = \frac{x_i - 3,32}{0,10}$$

وكلّ قيمة t_i تناسبها (في العامود (3)) قيمة معيّنة لوظيفة توزيع القانون الطبيعي المركز المختصر $\Pi(t_i)$. بالتالي ، احتمال أن ينتمي القطر لبرغي معيّن إلى الفئة (x_{i-1} – x_i) هو (العمود (4)) :

$$\begin{aligned} p_i = P\{x_{i-1} \leqslant X < x_i\} &= P\{X < x_i\} - P\{X < x_{i-1}\} \\ &= P\{T < t_i\} - P\{T < t_{i-1}\} \\ &= \Pi(t_i) - \Pi(t_{i-1}) . \end{aligned}$$



الشكل 30 . المدرج التكراري لتوزيع الأقطار



الشكل 31 . المدرج التكراري والمنحنى الطبيعي عند التسوية

إذن المقدار المسوّى لكلّ فئة (العمود (5)) يساوي np_i ، حيث n هي حجم الكمّية .

لنحسب مثلاً المقدار النظري للفئة 3,20 – 3,25 mm .

تبعاً للفرضية التي تقول أنّ قطر البراغي X موزّع حسب قانون طبيعي متغيّراه الوسيطيان $m = 3,32$ و$\sigma = 0.1$ ، احتمال أن ينتمي البرغي إلى هذه الفئة هو :

$$p_5 = P\{3{,}20 \leqslant X < 3{,}25\} = P\{X < 3{,}25\} - P\{X < 3{,}20\}$$
$$= P\{T < -0{,}7\} - P\{T < -1{,}2\}$$
$$= \Pi(-0{,}7) - \Pi(-1{,}2)\,.$$

الجدول 7 . حساب متوسّط توزيع الأقطار وانحرافه النموذجي (القراءة من اليسار إلى اليمين) :

الفئة	المقادير n_i	مركز الفئة x_i	المتغيّرة المساعدة x'_i	$n_i x'_i$	$n_i x'^2_i$
3,00 - 3,05	3	3,025	− 6	− 18	108
3,05 - 3,10	6	3,075	− 5	− 30	150
3,10 - 3,15	13	3,125	− 4	− 52	208
3,15 - 3,20	23	3,175	− 3	− 69	207
3,20 - 3,25	39	3,225	− 2	− 78	156
3,25 - 3,30	78	3,275	− 1	− 78	78
				− 325	
3,30 - 3,35	91	3,325	0	0	0
3,35 - 3,40	72	3,375	+ 1	+ 72	72
3,40 - 3,45	42	3,425	+ 2	+ 84	168
3,45 - 3,50	17	3,475	+ 3	+ 51	153
3,50 - 3,55	9	3,525	+ 4	+ 36	144
3,55 - 3,60	5	3,575	+ 5	+ 25	125
3,60 - 3,65	2	3,625	+ 6	+ 12	72
				+ 280	
المجموع	400	—	—	− 45	1 641

استبدال المتغيّرة :

$$x'_i = \frac{x_i - 3{,}325}{0{,}05}$$

حساب $\bar{x}$، σ_x :

$$\bar{x}' = \frac{-45}{400} = -0{,}112\,5$$

$$\sigma^2_{x'} = \frac{\sum n_i x'^2_i - n\bar{x}'^2}{n}$$

$$= \frac{1\,641 - 5{,}062\,5}{400} = 4{,}089\,844$$

$$\sigma_{x'} = \sqrt{4{,}089\,844} = 2{,}022$$

$$\bar{x} = 0{,}05\,\bar{x}' + 3{,}325$$
$$= -0{,}112\,5 \times 0{,}05 + 3{,}325$$
$$= 3{,}319 \approx 3{,}32$$

$$\sigma_x = 0{,}05\,\sigma_{x'}$$
$$= 0{,}05 \cdot 2{,}022$$
$$= 0{,}101 \approx 0{,}10$$

وذلك لأنّ القيمتين 3,25 = x_i و 3,20 = x_{i-1} تناسبهما :

$$t_i = \frac{3,25 - 3,32}{0,10} = -0,7 \quad \text{و} \quad t_{i-1} = \frac{3,20 - 3,32}{0,10} = -1,2 .$$

الجدول 8 . حساب المقادير المسوّاة . مقارنة مع المقادير الملحوظة . (القراءة من اليسار إلى اليمين) .

(1) الفئات $(x_{i-1}-x_i)$	(2) $t_i = \frac{x_i - 3,32}{0,10}$	(3) $\Pi(t_i)$	(4) الاحتمالات المسوّاة $p_i = \Pi(t_i) - \Pi(t_{i-1})$	(5) المقادير المسوّاة np_i	(6) المقادير الملحوظة n_i
		0,000 0			
3,00-3,05			0,003 5	1,4	3
	− 2,7	0,003 5			
3,05-3,10			0,010 4	4,2	6
	− 2,2	0,013 9			
3,10-3,15			0,030 7	12,2	13
	− 1,7	0,044 6			
3,15-3,20			0,070 5	28,2	23
	− 1,2	0,115 1			
3,20-3,25			0,126 9	50,8	39
	− 0,7	0,242 0			
3,25-3,30			0,178 7	71,5	78
	− 0,2	0,420 7			
3,30-3,35			0,197 2	78,9	91
	+ 0,3	0,617 9			
3,35-3,40			0,170 2	68,0	72
	+ 0,8	0,788 1			
3,40-3,45			0,115 1	46,1	42
	+ 1,3	0,903 2			
3,45-3,50			0,060 9	24,3	17
	+ 1,8	0,964 1			
3,50-3,55			0,025 2	10,1	9
	+ 2,3	0,989 3			
3,55-3,60			0,008 1	3,3	5
	+ 2,8	0,997 4			
3,60-3,65			0,002 6	1,0	2
		1,000 0			
المجموع	—	—	—	400,0	400

$$\Pi(-0{,}7) = 1 - \Pi(0{,}7) = 1 - 0{,}758\,0 = 0{,}242\,0$$
$$\Pi(-1{,}2) = 1 - \Pi(1{,}2) = 1 - 0{,}884\,9 = 0{,}115\,1 \,,$$

إذن :

$$p_3 = 0{,}242\,0 - 0{,}115\,1 = 0{,}126\,9$$

و :

$$np_3 = 400 \times 0{,}126\,9 = 50{,}8 \,.$$

ولكنّنا في الحقيقة لم نلاحظ لهذه الفئة سوى مقدار يساوي 39 . هل يمكننا إرجاع هذا الفارق وكذلك الفوارق الناجمة بالنسبة لبقيّة الفئات (الشكل 31) إلى التقلّبات العشوائية فقط ، أم أنّه يشكّك في صحّة التسوية ؟ هذا هو السؤال الذي سنحاول الإجابة عنه في القسم III .

B . التسوية البيانية : خط هنري

هناك طريقة بيانية (خطّية) لتسوية قانون طبيعي مع توزيع ملحوظ ، وتمثّل هذه الطريقة فائدة مزدوجة :

- فهي تسمح بتقييم الميزة الطبيعية للتوزيع الملحوظ على وجه التقريب ، أفضل ممّا على المدرج التكراري ؛
- وهي تعطي تقديراً بيانياً لمتوسّط التوزيع وانحرافه النموذجي .

خط هنري (Henri)

لنفترض أنّ الظاهرة المدروسة تتبع قانوناً طبيعياً . في هذه الحالة ، تكون الترددات المتراكمة الملحوظة على التوزيع مساوية تقريباً لقيم وظيفة توزيع القانون الطبيعي المناسبة :

$$F(x) = \Pi\left(\frac{x - m}{\sigma}\right) = \Pi(t) \,,$$

ويوجد بين قيمة المتغيّرة الإحصائية x والقيمة t المطابقة العلاقة التالية :

$$t = \frac{x - m}{\sigma} = \frac{1}{\sigma}x - \frac{m}{\sigma} \,,$$

وهي معادلة خطّ مستقيم .

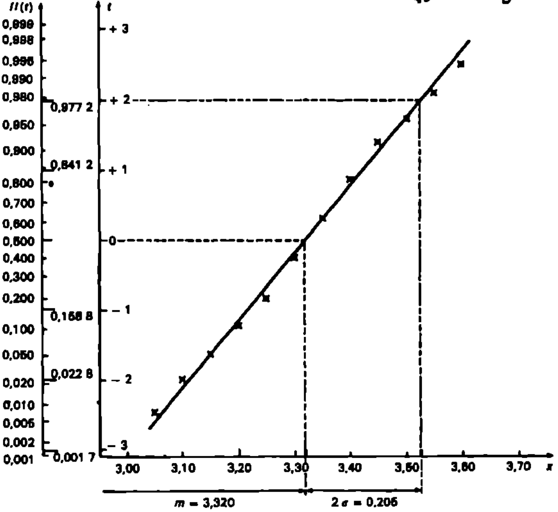
يمكننا إذن التأكّد بيانياً من طبيعية التوزيع . إذ بوسعنا انطلاقاً من التردّدات المتراكمة الملحوظة وبرجوعنا إلى جدول وظيفة توزيع القانون الطبيعي Π(t) أن نحدّد قيم t المناسبة . فإذا كان التوزيع الملحوظ يتبع فعلاً قانوناً طبيعياً، يجب أن تكون النقاط الحاصلة عند نقلنا قيم x وt على رسم بياني تقريباً على نفس الخط المستقيم .

مثلاً . لنعد إلى توزيع كمية من 400 برغي حسب أقطارها الذي سبق أن درسناه . يقدّم الجدول 9 حساب التردّدات المتراكمة Fi المناسبة لأطراف الفئات xi على الرسم البياني 32 وهي تبدو تقريباً على نفس الخطّ المستقيم : يمكننا إذن اعتبار التوزيع توزيعاً طبيعياً .

تحديد متوسّط التوزيع المسوّى m وانحرافه النموذجي σ بيانياً .

معادلة خطّ هنري المستقيم هي : $t = \frac{x - m}{\sigma}$

الشكل 32 . خطّ هنري



بالتالي :

ـ إذا كانت t = 0 ، إذن : x = m ؛

ـ إذا كانت t = 2 ، إذن : $x = m + 2\sigma$

الجدول 9 . خطّ هنري . حساب التردّدات المتراكمة وتحديد قيم t المناسبة (القراءة من اليسار إلى اليمين) :

الفئات $(x_{i-1}\text{-}x_i)$	المقادير n_i	المقادير المتراكمة N_i	التردّدات المتراكمة F_i	t_i
3,00-3,05	3	3	0,007 5	− 2,43
3,05-3,10	6	9	0,022 5	− 2,00
3,10-3,15	13	22	0,055 0	− 1,60
3,15-3,20	23	45	0,112 5	− 1,21
3,20-3,25	39	84	0,210 0	− 0,81
3,25-3,30	78	162	0,405 0	− 0,24
3,30-3,35	91	253	0,632 5	+ 0,34
3,35-3,40	72	325	0,812 5	+ 0,89
3,40-3,45	42	367	0,917 5	+ 1,39
3,45-3,50	17	384	0,960 0	+ 1,75
3,50-3,55	9	393	0,982 5	+ 2,11
3,55-3,60	5	398	0,995 0	+ 2,58
3,60-3,65	2	400	1,000 0	+ ∞
	400	—	—	—

وانطلاقاً من قراءة هذين الأمرين ، يمكننا تحديد قيمتي m و σ . هكذا نقرأ في المثل المعروض (الشكل 32) :

$$m = 3{,}320\,, \qquad m + 2\sigma = 3{,}525\,,$$

إذن :

$$2\sigma = 0{,}206 \qquad \sigma = 0{,}103 \approx 0{,}10$$

الرسم البياني القومي الحسابي (gausso-arithmétique)

التسوية البيانية أسهل للتطبيق من التسوية التحليلية : يمكننا أيضاً اختصارها .

كي نتجنّب البحث عن قيم t المناسبة الطرف كل فئة ، نستعمل على محور الإحداثيات الصادية مقياساً (سلّماً) غوسيّاً . هذا المقياس الوظيفي متدرّج ، تبعاً لطريقة شبيهة بالتي استخدمناها لبناء مقياس لوغارتمي (أنظر كتاب « الإحصاء الوصفي » الفصل IV) ، وذلك بنقلنا القيمة $\Pi(t)$ مقابل النقطة التي تبعد المسافة t عن مركز الانطلاق . ويُقام التناسب بين القيم المدوّرة لوظيفة التوزيع وقيم t انطلاقاً من الجدول $P(t)$ (أنظر ص 120) :

$\Pi(t)$	$P(t)$	
0,005	0,01	− 2,575 8
0,02	0,04	− 2,053 7
0,05	0,10	− 1,644 9
0,10	0,20	− 1,281 6
0,30	0,60	− 0,524 4
0,50	1	0
0,70	0,60	+ 0,524 4
0,90	0,20	+ 1,281 6
0,95	0,10	+ 1,644 9
0,98	0,04	+ 2,053 7
0,995	0,01	+ 2,575 8

ويسمح لنا تدرّج المقياس ببناء خطّ هنري مباشرة انطلاقاً من الترددات المتراكمة ، دون ضرورة لحساب قيم t المناسبة .

على الصعيد العملي ، نستعمل الأوراق القوميّة الحسابية الوظيفية : تتضمّن هذه الأوراق مقياساً غوسيّاً على أحد المحورين ومقياساً حسابياً على المحور الآخر (الشكل 33) . انطلاقاً من خطّ هنري المرسوم مباشرة على هذا الورق ، يمكننا بسهولة تحديد قيمتي المتوسّط m والانحراف النموذجي σ :

- إذا كانت $\Pi(t) = 0{,}5$ ، إذن $t = 0$ و $x = m$

- إذا كانت $\Pi(t) = 0{,}9772$ ، إذن $t=2$ و $x= m + 2\sigma$. ومن هنا نستنتج قيمتي m و σ .

0,998 0,995 0,990 0,980 0,977 2 0,950 0,900 0,800 0,700 0,600 0,500 0,400 0,300 0,200 0,100 0,050

(المقياس المئوي)

6 . قانون مشتق : القانون اللوغ ـ طبيعي

تتبع متغيّرة عشوائية معيّنة قانوناً لوغ ـ طبيعياً إذا كان لوغاريتمها (خوارزميتها) يتبع قانوناً طبيعياً .

ينتشر هذا القانون خاصة في مجال الظواهر الإجتماعية ـ الاقتصادية . في الواقع ، كل مرّة تكون فيها أسباب التغيّر ، التي توافق من ناحية أخرى شروط تطبيق نظرية الحدّ المركزي ، قابلة للضرب بعضها ببعض ، وليس للجمع ، تميل الظاهرة الملحوظة إلى أن تتبع قانوناً لوغ ـ طبيعياً .

A . قانون الاحتمال

في ما يلي ، وعدا تذكير معاكس محدّد بوضوح ، سنميّز متغيّرة لوغ ـ طبيعية X بواسطة المتغيّرين الوسيطيّين m و σ للقانون الطبيعي الذي يتبعه لوغاريتمها النيبري[1] (ln) :

$$\ln X = \mathcal{N}(m, \sigma) ;$$

m و σ هما إذن على التوالي أمل لوغاريتم X النيبري وانحرافه النموذجي :

$$T = \frac{\ln X - m}{\sigma}$$

تتبع قانوناً طبيعياً ممركزاً مختصراً . بالتالي يكون لدينا : $\ln X = m + t\sigma ,$

أي : $X = e^{m+t\sigma}$.

وبما أنّه لا يمكن تحديد اللوغاريتم إلاّ من أجل قيم المتغيّرة الإيجابية ، تتغيّر X من صفر إلى $+\infty$ فيما يتغيّر كلّ من ln X وT من ∞ إلى $+\infty$.

وظيفة التوزيع هي :

$$F(x) = \Pi\left(\frac{\ln x - m}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\ln x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right] d \ln x$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma}\int_{-\infty}^{\ln x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right] \frac{dx}{x} .$$

---

(1) بما أنّ اللوغاريتمات ذات القواعد المختلفة هي تناسبية ، إذا كان لوغاريتم X النيبري يتبع قانوناً طبيعياً ، فكذلك كل اللوغاريتمات الأخرى ، بصورة خاصّة اللوغاريتم العشري . استعمال اللوغاريتم النيبري يسهّل الحسابات .

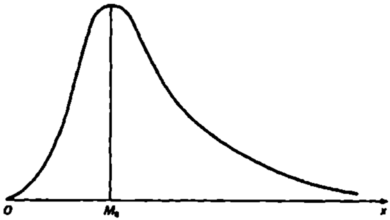
لأنّ :

$$d \ln x = \frac{dx}{x}$$

كثافة الاحتمال هي إذن (الشكل 34) :

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right]$$ ، إذا كانت $x > 0$

$= 0$ ، إذا كانت $x < 0$



الشكل 34 . شكل القانون اللوغ ـ طبيعي

القانون اللوغ ـ طبيعي هو توزيع غير متناظر ينبسط نحو اليمين .

B . مقاييس القانون اللوغ ـ طبيعي

مقاييس المتغيّرة اللوغ ـ طبيعية X ، ذات المتغيّرين الوسيطيّين m وσ، هي :

المنوال : $M_0 = e^{m-\sigma^2}$

الأمل الرياضي : $E\{X\} = e^{m+\sigma^2/2}$

التباين : $V\{X\} = e^{2(m+\sigma^2)}(1 - e^{-\sigma^2})$.

وتُستنتج هذه العبارات مباشرة من تطبيق قواعد التعريف .

مثلاً : حساب الأمل الرياضي . انطلاقاً من التعريف :

$$E\{X\} = \frac{1}{\sqrt{2\pi}\,\sigma} \int_0^\infty x \exp\left[-\frac{1}{2}\left(\frac{\ln x - m}{\sigma}\right)^2\right] \frac{dx}{x}$$

لنضع :

$$t = \frac{\ln x - m}{\sigma}$$

بالتالي :

$$x = e^{m+\sigma t}, \qquad \frac{dx}{x} = \sigma\, dt$$

فيصبح لدينا :

$$E\{X\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{m+\sigma t}\, e^{-t^2/2}\, dt = e^{m+\sigma^2/2} \times \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-(t-\sigma)^2/2}\, dt\,.$$

وإذا وضعنا $u = t - \sigma$.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-(t-\sigma)^2/2}\, dt = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-u^2/2}\, du = 1\,,$$

لأنّ هذا التكامل يبدو وكأنّه مجموع احتمالات المتغيّرة الطبيعية المركزة المختصرة u .

بالتالي :

$$E\{X\} = e^{m+\sigma^2/2}\,.$$

C . تحديد الاحتمالات عملياً

تُحدّد احتمالات المتغيّرة اللوغ ـ طبيعية X انطلاقاً من جداول القانون الطبيعي المركز المختصر .

في الواقع ، يسمح لنا استبدال المتغيّرة التالي :

$$t = \frac{\log x - m}{\sigma}$$

حيث log x تعني لوغاريتم (خوارزمية) x ، بالبحث في الجدول $\Pi(t)$ عن الاحتمال أن تكون المتغيّرة اللوغ ـ طبيعية X أصغر من قيمة معطيّة x :

$$P\{X < x\} = P\{T < t\} = \Pi(t)\,.$$

مثل 1 . لنفترض X متغيّرة لوغ ـ طبيعية يتبع لوغاريتمها العشري قانوناً طبيعياً متغيّراه الوسيطيان هما m=3 و σ=0,2 . ما هو احتمال أن تكون X أصغر من 7500 ؟ .

قيمة المتغيّرة الطبيعية المركزة المختصرة التي تطابق X =7500 هي :

$$t = \frac{\log_{10} 7\,500 - 3}{0{,}2}, \qquad t = \frac{3{,}875\,06 - 3}{0{,}2} = 0{,}437\,53$$

$$P\{X < 7\,500\} = P\{T < 0{,}487\,53\} = 0{,}6691\,,$$

وذلك بواسطة استكمال في الجدول $\Pi(t)$.

وبالعكس ، من الممكن تحديد قيمة معيّنة x نعرف قيمة احتمال أن تكون X أصغر منها.

مثل 7 . لنفترض X متغيّرة لوغ - طبيعية يتبع لوغاريتمها النبيري lnX قانوناً طبيعياً بمتغيّرين وسيطين m=2 و0,4= σ . ما هي قيمة وسيط وربيعي التوزيع[1] :

$$t = \frac{\ln x - m}{\sigma}$$

إذن :

$$\ln x = m + t\sigma\,, \qquad x = e^{m+t\sigma}$$

لنحسب قيم t المناسبة للوسيط وللربيعين .

بالنسبة للوسيط M :

$P\{T < t\} = 0{,}5$ إذن : $t = 0$

بالنسبة للربيع الأول Q_1 :

$$P\{T < t\} = 0{,}25\,,$$

إذن $t = -0{,}6745$ بعد استعانتنا بالجدول P(t) .

بالنسبة للربيع الثالث Q_3 ، لدينا بحكم التناظر :

$P\{T < t\} = 0{,}75\,,$ إذن $t = +0{,}674\,5\,.$

بالتالي (بعد وضع t بقيمتها في كلّ من الحالات الثلاث) :

$$M = e^m\,,$$
$$Q_1 = e^{m-0{,}6745\sigma}\,,$$
$$Q_3 = e^{m+0{,}6745\sigma}\,.$$

عددياً يمكن إجراء الحساب بواسطة اللوغاريتمات العشرية :

$$\log_{10} M = m \log_{10} e$$
$$= 2 \times 0{,}434\,29 = 0{,}868\,58\,.$$

---

(1) انظر المجلّد الأول : « الإحصاء الوصفي » ، الفصل VI .

إذن : $M = 7,39$;

$$\log_{10} Q_1 = (m - 0,674\,5\,\sigma).\log_{10} e$$
$$= (2 - 0,269\,8) \times 0,434\,29 = 0,751\,41 ,$$

إذن : $Q_1 = 5,64$;

$$\log_{10} Q_3 = (m + 0,674\,5\,\sigma).\log_{10} e$$
$$= (2 + 0,269\,8) \times 0,434\,29 = 0,985\,75 ,$$

إذن : $Q_3 = 9,68$

D . شروط التطبيق

تنتج شروط تطبيق القانون اللوغ ـ طبيعي عن الشروط التي وضعناها للقانون الطبيعي (أنظر ص 110) : يكفي بالواقع أن يفي logX بمتطلّبات صحّة نظرية الحدّ المركزي . وهذا يتحقق عندما تكون X نتيجة عدد كبير من العوامل المستقلّة X_i ، يكون وزن كلّ منها صغيراً جداً بالنسبة للمجموعة ، وتتآلف تأثيراتها الإيجابية فيما بينها بالضرب ، وليس بالجمع كما في حالة القانون الطبيعي :

$$X = X_1.X_2.\dots.X_n = \prod_{i=1}^{n} X_i .$$

وإذا أخذنا لوغاريتم X ، تتحوّل هذه العبارة بالفعل إلى سلسلة عوامل ممكنة الجمع $\log X_i$:

$$\log X = \log X_1 + \log X_2 + \cdots + \log X_n = \sum_{i=1}^{n} \log X_i ,$$

وهذه السلسلة تفي بشروط صحّة نظرية الحدّ المركزي ، ما يعني أنّ log X يميل نحو القانون الطبيعي عندما تتزايد n بصورة غير متناهية .

إلاّ أنّه في المجال الاقتصادي والاجتماعي حيث نلتقي القانون اللوغ طبيعي باستمرار (توزيعات الرواتب ، توزيعات أرقام المبيعات ، وبشكل عام توزيعات الوحدات الاقتصادية حسب أحجامها) ، ليس من الممكن دوما تبرير استعمال هذا القانون بواسطة اعتبارات نظرية . يجب إذن النظر إليه كمجرّد نموذج وصفي يطابق الظاهرة وليس له أي قيمة تفسيرية .

E . تسوية قانون لوغ ـ طبيعي مع توزيع إحصائي ملحوظ

تجري التسوية حسب طرق شبيهة بالطرق المستعملة للقانون الطبيعي . يمكننا ،

بشكل خاص ، اعتماد تسوية بيانية على طريقة خطّ هنري . في هذه الحالة ننقل قيم log x على محور الإحداثيات السينيات وقيم T المطابقة على محور الاحداثيات الصاديات . وكي نتجنب البحث عن هذه القيم في الجداول ، نستعمل عادة الأوراق الغوسّية ـ اللوغاريتمية التي تباع في المكتبات ، والتي يكون محور إحداثياتها السينيات مدرّجاً حسب قياس لوغاريتمي ومحور احداثياتها الصاديات حسب مقياس غوسيّ .

F . تعميم القانون اللوغ ـ طبيعي

لنفترض X متغيّرة لوغ ـ طبيعية . بمتغيّرين وسيطيّين m و σ: لنحدّد المتغيّرة Y بواسطة استبدال مركز الانطلاق :

$$Y = X + x_0$$

بالتالي يتبع $\log (Y - x_0)$ قانوناً طبيعياً بمتغيّرين وسيطيّين m و σ :

$$\frac{\log(Y - x_0) - m}{\sigma} = T = \mathcal{N}(0,1) .$$

وتُسمّى المتغيّرة Y متغيّرة لوغ ـ طبيعية معمّمة . وهي تتعلّق بثلاثة متغيّرات وسيطية : m, x_0 و σ .

القسم II
قانون χ^2

1 . تعريف .ـ 2 .ـ المقاييس : A . الأمل الرياضي ؛ B . التباين .ـ 43 . شروط التطبيق : A . عدد درجات الحرية ؛ B . مجموع متغيّرات χ^2 مستقلّة .ـ 4 . جدول قانون χ^2 .

إن قانون كي اثنان أو مربّع كي χ^2 هو قانون مهمّ ، لا لتمثيل سلاسل إحصائية ملحوظة كما في حالة القوانين التي درسناها ولكن بحكم الدور الذي يلعبه في الاختبارات الإحصائية ، بصورة خاصّة اختبار تسوية قانون نظري مع توزيع ملحوظ (أنظر القسم III) .

1 . تعريف

قانون الاحتمال

لنفترض $T_1, T_2, ..., T_\nu$ متغيّرةً عشوائية طبيعية ممركزة مختصرة ومستقلّة . إنّ مجموع مربّعاتها هو أيضاً متغيّرة عشوائية جرت العادة على الإشارة إليه بواسطة الحرف

اليوناني كي (Khi) مرفوعاً إلى مربّعه (إشارة أدرجها K. Pearson, 1905 بيرسون) :

$$\chi^2 = T_1^2 + T_2^2 + \cdots + T_\nu^2 .$$

وتتغيّر هذه المتغيّرة العشوائية بين صفر وما لا نهاية ، وكثافة احتمالها هي

$$f(\chi^2) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\,\chi^{\nu-2}\,e^{-\chi^2/2} .$$

ونقول أنها تتبع قانون χ^2 ذا ν درجة حرّية . أمّا $2^{\nu/2}\,\Gamma(\nu/2)$ فهي ثابتة بحيث يساوي مجموع الاحتمالات واحداً :

$$\int_0^{+\infty} f(\chi^2)\, d(\chi^2) = 1$$

ملحق رياضيات

دالّة أولر (Euler) من النوع الثاني

$\Gamma(n)$ هي دالّة أولر من النوع الثاني (الدّالة غمّا gamma) ، وهي محدّدة بواسطة التكامل .

$$\Gamma(n) = \int_0^{+\infty} e^{-x}\, x^{n-1}\, dx ,$$

حيث n هو متغيّر وسيطي إيجابي .

إذا اعتمدنا التكامل بالتجزئة :

$$\int u\, dv = uv - \int v\, du ,$$

$$u = x^{n-1} , \qquad dv = e^{-x}\, dx ,$$

$$du = (n-1)\, x^{n-2} , \qquad v = -\, e^{-x} ,$$

$$\Gamma(n) = [-\, x^{n-1}\, e^{-x}]_0^{+\infty} + (n-1)\int_0^{+\infty} e^{-x}\, x^{n-2}\, dx .$$

القسم الأوّل من المجموع يساوي صفراً فيما يساوي القسم الثاني ، انطلاقاً من التعريف : $(n-1)\,\Gamma(n-1)$:

$$\Gamma(n) = (n-1)\,\Gamma(n-1) ,$$

إذن بالتكرار :

$$\begin{aligned}\Gamma(n) &= (n-1)\,\Gamma(n-1)\\ \Gamma(n-1) &= (n-2)\,\Gamma(n-2)\\ &\cdots\cdots\\ \Gamma(n-k+1) &= (n-k)\,\Gamma(n-k)\\ \hline \Gamma(n) &= (n-1)(n-2)\ldots(n-k)\,\Gamma(n-k)\end{aligned}$$

حيث n هو عدد إيجابي و k عدد صحيح ، كي نحسب قيمة $\Gamma(n)$ يكفي أن نعرف قيم $\Gamma(n-k)$ عندما $0 < n-k < 1$.

قيمة $\Gamma(n)$ تساوي 1 :

$$\Gamma(1) = \int_0^{+\infty} e^{-x}\,dx = 1$$

بالتالي ، عندما يكون n صحيحاً :

$$\Gamma(n) = (n-1)(n-2)\ldots 1 . \Gamma(1) = (n-1)!$$

تحقق الدّالة Γ استكمال الدّالة العاملية .

هناك قيمة مهمّة جداً خاصة لدراسة قانون π^2 ، وهي $\Gamma(1/2)$:

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,.$$

بالفعل ، انطلاقاً من التعريف :

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} e^{-x}\,x^{-1/2}\,dx\,.$$

لنضع $x = t^2/2,\ dx = t\,dt$:

$$\Gamma\left(\frac{1}{2}\right) = \int_0^{+\infty} e^{-t^2/2}\left(\frac{t^2}{2}\right)^{-1/2} t\,dt = \sqrt{2}\int_0^{+\infty} e^{-t^2/2}\,dt\,.$$

لكن $\int_0^{+\infty} e^{-t^2/2}\,dt$ يساوي $\frac{\sqrt{2\pi}}{2} = \frac{\sqrt{\pi}}{\sqrt{2}}$ (إذ أنّ $\frac{1}{\sqrt{2\pi}}\int_0^{+\infty} e^{-t^2/2}\,dt$

وهو تكامل القانون الطبيعي المركز المختصر ، مأخوذاً بين صفر و $+\infty$ ، يساوي 1/2 .
بالتالي :

$$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\,.$$

إذن ، إذا كانت n عدداً مفرداً :

$$\Gamma\left(\frac{n}{2}\right) = \left(\frac{n}{2} - 1\right)\left(\frac{n}{2} - 2\right) \dots \frac{1}{2}\,\Gamma\left(\frac{1}{2}\right) = \left(\frac{n}{2} - 1\right)\left(\frac{n}{2} - 2\right) \dots \frac{1}{2}\sqrt{\pi}.$$

قانون 8 أو قانون بيرسون Pearson من النوع III

الاحتمال النموذجي لقانون χ^2 ذي ν درجة حرّية هو :

$$f(\chi^2)\, d(\chi^2) = \frac{1}{2^{\nu/2}\,\Gamma(\nu/2)}\, \chi^{\nu-2}\, e^{-\chi^2/2}\, d(\chi^2) \qquad (1)$$

لنضع :

$$x = \frac{\chi^2}{2} \qquad \text{إذن} \qquad dx = \frac{d(\chi^2)}{2}$$

يمكن أن نكتب المعادلة (1) على الشكل التالي :

$$f(\chi^2)\, d(\chi^2) = \frac{1}{\Gamma(\nu/2)}\, \frac{(\chi^2)^{\nu/2-1}}{2^{\nu/2-1}}\, e^{-\chi^2/2}\, \frac{d(\chi^2)}{2},$$

$$f(x)\, dx = \frac{1}{\Gamma(\nu/2)}\, x^{\nu/2-1}\, e^{-x}\, dx.$$

هذا هو الاحتمال النموذجي لقانون $\delta_{\nu/2}$ (قانون بيرسون من النوع III) بمتغيّر وسيطي $\nu/2$ ، الذي يعادل القانون χ^2 . يمكننا التحقّق من أنّ مجموع الاحتمالات يساوي 1 :

$$\int_0^\infty f(x)\, dx = \frac{1}{\Gamma(\nu/2)} \int_0^\infty x^{\nu/2-1}\, e^{-x}\, dx,$$

لأنّه ، انطلاقاً من التعريف :

$$\int_0^\infty x^{\nu/2-1}\, e^{-x}\, dx = \Gamma(\nu/2).$$

عدا عن الأهمّية التي يمثّلها القانون 8 لدراسة قانون χ^2، فهو يلعب ، بحدّ ذاته ، دوراً مهمّاً في دراسة سياقات بواسّون Poisson العشوائية (أنظر الفصل II ، القسم III ، ص 96) . إذا كان احتمال تحقيق حدث معيّن ، خلال فترة لا متناهية الصغر من الوقت dt ، يساوي pdt ، حيث تبقى p ثابتة طيلة فترة الملاحظة ، إذن :

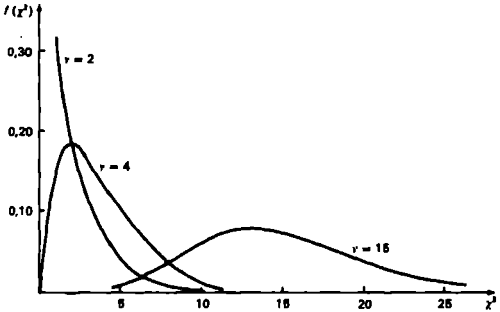
ـ قانون الحوادث التي تأتي أثناء فسحة من الوقت T هو قانون بواسّون بمتغيّر وسيطي pT ؛

ـ قانون فسحة الوقت التي تفصل بين ظهور حدثين متاليين هو قانون من النوع δ_1 .
ـ قانون فسحة الوقت التي تفصل بين أوّل وآخر حدث من سلسلة تتألّف من n حادثاً متالياً هو قانون من النوع δ_n .

للقانون δ إذن تطبيقات مهمّة في مجال صفوف الانتظار : انتظار زبائن على شباك معيّن ، سيارات في مركز للشحن ، آلات للتصليح ، الخ .

الشكل

توزيع χ^2 هو توزيع غير متناظر مع انبساط نحو اليمين . إلّا أنّه يميل الى أن يصبح متناظراً عندما يتزايد عدد درجات الحرّية ν : عندها يقترب من التوزيع الطبيعي ويمكننا مطابقته معه عندما يكون ν أكبر من 30 (الشكل 35) .



الشكل 35 . أشكال قانون χ^2 حسب عدد درجات الحرية ν

2 . مقاييس قانون χ^2

A . الأمل الرياضي

الأمل الرياضي (أو المعدّل الوسطي) لقانون χ^2 ذي ν درجة حرّية يساوي

$$E\{\chi^2\} = \nu .$$

البرهان : انطلاقاً من تعريف الأمل الرياضي :

$$E\{\chi^2\} = \int_0^{\infty} \chi^2 f(\chi^2)\, d(\chi^2) .$$

إذا أجرينا استبدال المتغيّرة التالي :

$$x = \frac{\chi^2}{2}$$

نحصل من جهة على :

$$f(\chi^2)\, d(\chi^2) = \frac{1}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x}\, dx$$ (أنظر ص 144)

ومن جهة أخرى :

$$\chi^2 = 2\,x$$

إذن :

$$E\{\chi^2\} = \frac{2}{\Gamma(\nu/2)} \int_0^{\infty} x^{\nu/2} e^{-x}\, dx .$$

ولكن التكامل يساوي

$$E\{\chi^2\} = 2\,\frac{\Gamma(\nu/2+1)}{\Gamma(\nu/2)} = 2\,\frac{\frac{\nu}{2}\Gamma(\nu/2)}{\Gamma(\nu/2)} = \nu .$$

B . التباين

تباين قانون χ^2 ذي ν درجة حرّية يساوي 2ν :

$$V\{\chi^2\} = 2\,\nu .$$

إذن تباين قانون χ^2 يساوي مضاعف معدّله الوسطي .

البرهان . يمكننا التعبير عن تباين متغيّرة عشوائية بواسطة أوّل عزمين (أنظر الفصل 1 ، ص 63) .

$$V\{X\} = E\{X^2\} - [E\{X\}]^2 ,$$
$$V\{\chi^2\} = E\{(\chi^2)^2\} - [E\{\chi^2\}]^2 .$$

انطلاقاً من تعريف الأمل الرياضي :

$$E\{(\chi^2)^2\} = \int_0^{\infty} (\chi^2)^2 f(\chi^2)\, d(\chi^2) .$$

إذا وضعنا : $x = \chi^2/2$ ، نحصل من جهة على :

$$f(\chi^2)\,d(\chi^2) = \frac{1}{\Gamma(\nu/2)} x^{\nu/2-1} e^{-x} dx ,$$

كما سبق أن رأينا ، ومن جهة أخرى :

$$(\chi^2)^2 = 4\,x^2 ,$$

إذن :

$$E\{(\chi^2)^2\} = \frac{4}{\Gamma(\nu/2)} \int_0^{\infty} x^{\nu/2+1} e^{-x} dx .$$

إلّا أنّ التكامل يساوي $\Gamma(\nu/2 + 2)$

$$E\{(\chi^2)^2\} = 4.\frac{\Gamma(\nu/2+2)}{\Gamma(\nu/2)} = 4.\frac{(\nu/2+1)\,\nu/2\,\Gamma(\nu/2)}{\Gamma(\nu/2)} = \nu^2 + 2\,\nu .$$

. بالتالي :

$$V\{\chi^2\} = E\{(\chi^2)^2\} - [E\{\chi^2\}]^2 = \nu^2 + 2\,\nu - \nu^2 = 2\,\nu .$$

3 . شروط تطبيق قانون χ^2

لقد حدّدنا المتغيرة العشوائية χ^2 بِـ ν درجة حرّية كمجموع مربّعات ν متغيّرة طبيعية ممركزة مختصرة مستقلّة :

$$\chi^2 = T_1^2 + T_2^2 + \cdots + T_\nu^2 .$$

بما أنّ المتغيّرات العشوائية $T_1, T_2, \ldots, T_\nu$ مستقلّة ، فإنّ احتمال أن توجد النقطة العشوائية $(T_1, T_2, \ldots, T_\nu)$ من الفضاء ذي الـ ν بعداً في عنصر حجم تفاضلي حول النقطة M ذات الإحداثيات $(t_1, t_2, \ldots, t_\nu)$ هو (قاعدة الاحتمالات المركّبة) :

$$P\{t_1 \leq T_1 < t_1+dt_1, t_2 \leq T_2 < t_2+dt_2, \ldots, t_\nu \leq T_\nu < t_\nu+dt_\nu\}$$
$$= f(t_1)\,dt_1 . f(t_2)\,dt_2 \ldots . f(t_\nu)\,dt_\nu = \frac{1}{(2\pi)^{\nu/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{\nu} t_i^2\right) dt_1\,dt_2 \ldots dt_\nu .$$

في هذا الفضاء ذي الـ ν بعداً : $\chi^2 = t_1^2 + t_2^2 + \ldots + t_\nu^2$ تمثّل مربّع المسافة من النقطة M إلى مركز الانطلاق . بصورة خاصة ، كلّ النقاط التي لها نفس كثافة الاحتمال :

$$\frac{1}{(2\pi)^{\nu/2}} \exp\left(-\frac{1}{2}\sum_{i=1}^{\nu} t_i^2\right)$$

توجد على سطح كرة مركزها نقطة الانطلاق وشعاعها :

$$\chi = \sqrt{t_1^2 + t_2^2 + \cdots + t_\nu^2}\,.$$

في نظام الإحداثيات الكرويّة الجديد هذا ، عنصر الحجم التفاضلي محصور بين كرتين شعاعهما χ و$\chi + d\chi$ مع فارق ثابتة :

$$\chi^{\nu-1}\, d\chi = (\chi^2)^{\nu/2-1}\, d(\chi^2)\,.$$

أخيراً عبارة احتمال المتغيرة العشوائية χ^2 هي :

$$f(\chi^2)\, d(\chi^2) = K\, e^{-\chi^2/2}\, (\chi^2)^{\nu/2-1}\, d(\chi^2)\,.$$

وتُحدّد الثابتة K بشكل يكون فيه مجموع الاحتمالات النموذجية مساوياً لواحد :

$$K \int_0^\infty e^{-\chi^2/2}\, (\chi^2)^{\nu/2-1}\, d(\chi^2)\,.$$

وكما رأينا :

$$K = \frac{1}{2^{\nu/2}\, \Gamma^{\nu/2}}\,.$$

A . عدد درجات الحرّية

أن نأخذ بعين الاعتبار ν متغيّرة طبيعية ممركزة مختصرة مستقلّة يعني أن نضع أنفسنا في فضاء ذي ν بعداً : عدد درجات حرّية قانون χ^2 الذي يتبعه مجموع مربّعات هذه المتغيّرات يطابق عدد أبعاد الفضاء الذي يحتوي النقاط التي تمثّل قيم χ^2 .

لنأخذ الآن n متغيّرة طبيعية مترابطة خطّياً : عندها يكون عدد أبعاد الفضاء الذي يتضمّن النقاط التمثيلية ν أصغر من n .

إذا وجد مثلاً بين المتغيّرات العلاقة الخطّية التالية :

$$a_1 T_1 + a_2 T_2 + \cdots + a_n T_n = a_0\,,$$

توجد النقاط التمثيلية في فضاء ذي $\nu = n - 1$ بعداً .

في حال وجود علاقتين خطّيتين ، يصبح عدد أبعاد الفضاء $\nu = n - 2$ ، الخ . بالتالي ، يكون لتوزيعات χ^2 المناسبة $\nu = n-1$ ، $\nu = n-2$ ، الخ درجة حرّية .

B مجموع متغيّرات χ^2

إنّ مجموع متغيّرتين χ^2 مستقلّتين لهما على التوالي ν_1 وν_2 درجة حرّية يتبع هو نفسه قانون χ^2 ذا درجة حرية $\nu = \nu_1 + \nu_2$.

بالطبع يمكننا بسط هذه النتيجة إلى أي عدد من متغيّرات χ^2 مستقلّة : إذا كانت $\chi_1^2, \chi_2^2, \dots, \chi_k^2$ ، k متغيّرة مستقلّة ذات $\nu_1, \nu_2, \dots, \nu_k$ درجة حرّية ، فإنّ مجموعها :

$$\chi_1^2 + \chi_2^2 + \cdots + \chi_k^2$$

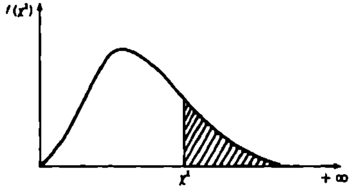
$$\nu = \nu_1 + \nu_2 + \cdots + \nu_k$$

درجة حرّية .

4 . جدول قانون χ^2

بما أنّ توزيع χ^2 لا يرتبط إلاّ بمتغيّر وسيطي واحد ν ، وهو عدد درجات الحرّية ، فإن للجدول الذي نجده في ملحق الكتاب (الجدول 5) مدخلاً مزدوجاً (ν و P) . وهو يعطي قيمة χ^2 التي يساوي احتمال تجاوزها P ، وذلك لقيم ν الأصغر من أو التي تساوي 30 (الشكل 36) :

$$P = \int_{\chi^2}^{\infty} f(\chi^2)\, d(\chi^2) .$$



الشكل 36 . تفسير جدول توزيع χ^2

مثلاً . إذا كانت $\nu = 7$ ، فإنّ احتمال أن تكون قيمة χ^2 أكبر من 2,83 هو 90% واحتمال أن تكون أكبر من 14,07 هو 5% .

إنّ وضع هذا الجدول ، كما سنرى في القسم الذي يلي ، هو متكيّف تماماً مع اختبار تسوية قانون نظري مع توزيع ملحوظ .

القسم III

صحّة تسوية قانون نظري مع توزيع ملحوظ

1 . تحديد وقانون احتمال المسافة بين توزيع ملحوظ والقانون النظري المناسب .ـ 2 . إختبار χ^2 . ـ 3 . أمثلة : اختبارات تسوية قانون ذي حدّين ، قانون بواسّون وقانون طبيعي .

لنفترض أنّ متغيّرة إحصائية معيّنة تتبع تماماً قانون احتمال معيّن P . إذا أخذنا عيّنة من المجتمع الإحصائي المطابق لهذا القانون ، فإنّ التوزيع الملحوظ سينحرف دوماً وبدرجة متفاوتة عن التوزيع النظري : إذ تكون الحالات الملحوظة مشوبة بتقلّبات عشوائيّة .

بشكل عام ، نجهل شكل قانون الاحتمال الذي تتبعه الظاهرة الملحوظة ، وكذلك قيمة متغيّرات هذا القانون الوسيطية . ونصل إلى اختيار قانون الاحتمال الذي يبدو مناسباً عبر تفكير حول طبيعة الظاهرة وتحليل التوزيع الملحوظ ، بعدها نقدّر متغيّرات هذا القانون انطلاقاً من السلسلة التجريبية .

يمكن إذن نسب الانحرافات بين القانون النظري المحدّد بهذه الطريقة والتوزيع الملحوظ :

- إمّا إلى تقلّبات المعاينة ،
- إمّا إلى كون الظاهرة لا تتبع في الحقيقة القانون المفترض .

بصورة أسهل : إذا كانت الانحرافات ضعيفة بما فيه الكفاية ، نسلّم بكونها عائدة إلى التقلّبات العشوائية ؛ أمّا إذا كانت مرتفعة ، نستنتج أنّه لا يمكن إلقاؤها على عاتق التقلّبات فقط وأنّ الظاهرة لا تتبع القانون المأخوذ .

بشكل أدقّ ، الحكم على صحّة تسوية معيّنة يعني أن نختبر الفرضية التي تقول بأنّ الظاهرة الملحوظة تتبع القانون النظري المفترض . للقيام بهذا الأمر يجب أوّلاً تحديد قياس للمسافة الموجودة بين التوزيع التجريبي وقانون الاحتمال النظري ، ثمّ تحديد قانون احتمال هذه الكمّية .

عند معرفتنا لهذا القانون ، إذا لاحظنا في الفرضية المأخوذة احتمالاً قوياً للحصول ، بحكم التقلّبات العشوائية فقط ، على مسافة أكبر من المسافة الملحوظة ، نقبل الفرضية ونسلّم بأنّ الظاهرة تتبع فعلاً القانون النظري المفترض ؛ أمّا إذا كان

هـذا الاحتمـال ضعيفـاً (أصغر من 5% مثـلاً) ، فهنـاك فـرص كبيـرة لأن تكـون الانحرافات الملحوظة غير عائدة إلى مجرّد التقلّبات العشوائية ، ولكن إلى عدم مـوافقة القانون النظري المأخوذ لتمثيل الظاهرة : عندها نرمي الفرضية .

1_1 . تحديد وقانون احتمال المسافة بين التوزيع الملحوظ والقانون النظري المناسب

لنفترض X متغيّـرة عشوائية تتبع قانون احتمال نظرياً P .

أن نجري N ملاحظة لهذه المتغيّـرة يعني أن نسحب عيّـنة حجمها N من المجتمع الإحصائي اللامتناهي الذي يطابق قانون الاحتمال P . تُنظَّـم الملاحظات حسب k كيفيّـة :

$$C_1, C_2, \dots, C_k$$

تمثّـل مختلف قيم المتغيّـرة الممكنة أو مجموعات قيمها إذا كانت متغيّـرة منفصلة ، أو فئات قيم المتغيّـرة إذا كانت متواصلة .

لكلّ من هذه الكيفيات أو الفئات احتمال يحدّده القانون P :

$$p_1, p_2, \dots, p_k .$$

والمقدار الذي يمكننا ملاحظته على العيّـنة لكلّ من هذه الفئات :

$$\xi_1, \xi_2, \dots, \xi_k$$

هو متغيّـرة عشوائية ذات حدّين .

هكـذا ، بالنسبـة للفئة C_i ، المقـدار ξ_i هـو متغيّـرة ذات حـدّين بمتغيّـرين وسيطيّين N ، مقدار العيّـنة ، وp_i ، احتمال أن تنتمي المتغيّـرة X إلى هذه الفئة :

$$\xi_i = \mathscr{B}(N, p_i) .$$

أمله الرياضي :

$$E\{\xi_i\} = Np_i$$

يمثّل المقدار النظري للفئة C_i .

وتغيّـره هو :

$$V\{\xi_i\} = Np_i(1 - p_i) \approx Np_i .$$

في الواقع يجري اختيار عدد الفئات وحدودها بشكل يكون فيه الاحتمال p_i صغيراً نسبياً ، إذاً تكون الكمّية $1 - p_i$ قريبة من 1 .

في هذه الشروط ، وعلى أساس أن تكون الفئة C_i كبيرة بما فيه الكفاية للحصول على مقدار نظري يساوي على الأقلّ 4 أو 5 وحدات إحصائية (وإلّا تبقى شروط ميل القانون ذي الحدّين نحو القانون الطبيعي ناقصة) ، يمكن اعتبار الانحراف المختصر E_i بين المقدار التجريبي والمقدار النظري :

$$E_i = \frac{\xi_i - Np_i}{\sqrt{Np_i}}$$

متغيّرة طبيعية ممركزة مختصرة .

المقادير الملحوظة حقيقة على العيّنة لكلّ من الفئات هي :

$$N_1, N_2, \ldots, N_k .$$

هكذا ، بالنسبة للفئة C_i ، يأخذ الانحراف المختصر على العيّنة القيمة :

$$e_i = \frac{N_i - Np_i}{\sqrt{Np_i}} .$$

لنرفع كلّ هذه الانحرافات إلى مربّعاتها ونأخذ مجموعها لكلّ الفئات :

$$d = \sum_{i=1}^{k} e_i^2 = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i} .$$

يقدّم هذا المجموع d قياساً للمسافة الموجودة بين التوزيع الملحوظ والتوزيع النظري .

ونعرف أنّ المتغيّرة العشوائية :

$$D = \sum_{i=1}^{k} E_i^2 = \sum_{i=1}^{k} \frac{(\xi_i - Np_i)^2}{Np_i}$$

التي تمثّل d قيمتها الملحوظة على العيّنة ، هي مجموع مربّعات k متغيّرة طبيعية ممركزة مختصرة ترتبط في ما بينها العلاقة الخطّية التالية :

$$\xi_1 + \xi_2 + \cdots + \xi_k = N .$$

إذن ، تتبع هذه المتغيّرة قانون χ^2 ذا $\nu = k-1$ درجة حرية (أنظر القسم II) .

وهذه الميزة جديرة بالملاحظة : في الواقع لا يتوقف قانون احتمال D إلا على عدد الفئات ، وليس على طبيعة الظاهرة موضع الدراسة (أي قانون الاحتمال P) .

2 . اختبار χ^2

بشكل عام ، لا نعرف مسبقاً قانون الاحتمال النظري الذي تتبعه المتغيّرة العشوائية X . حسب طبيعة الظاهرة وبعد تحليل التوزيع الملحوظ ، نختار نموذج قانون نقدّر متغيّراته الوسيطية على أساس الحالات الملحوظة (أنظر : القانون ذو الحدّين ، ص 78 ؛ قانون بواسّون ، ص 98 ؛ القانون الطبيعي ، ص 126) . هذا القانون المفترض P يعطي للكيفيات أو الفئات k التالية :

$$C_1, C_2, \ldots, C_k ,$$

الاحتمالات : $p_1, p_2, \ldots, p_k$ وكذلك المقادير النظرية : $Np_1, Np_2, \ldots, Np_k$
وهكذا يمكننا حساب القيمة :

$$d = \sum_{i=1}^{k} \frac{(N_i - Np_i)^2}{Np_i}$$

التي تأخذها المتغيّرة العشوائية D ، التي تقيس المسافة الموجودة بين التوزيع الملحوظ والتوزيع النظري .

وتتبع هذه المسافة D ، حسب الفرضية حيث التوزيع النظري هو فعلاً القانون P ، قانون χ^2 . ويتوقف عدد درجات حرّية هذا القانون على عدد الفئات k وعدد المتغيّرات الوسيطية المقدّرة r انطلاقاً من الحالات الملحوظة :

$$\nu = k - r - 1 .$$

في الواقع ، عدا عن العلاقة : $N_1 + N_2 + \ldots + N_k = N$

يوجد بين المقادير الملحوظة N_i في كلّ فئة علاقة أو عدّة علاقات إضافية يوجدها تقدير متغيّر وسيطي أو أكثر .

في حالة القانون ذي الحدّين الذي نقدّر متغيّره الوسيطي p بواسطة $\bar{x}/n$ حيث $\bar{x}$ هي المعدّل الوسيطي الملحوظ و n متغيّر القانون الوسيطي الثاني (أنظر المثل 1) ، لدينا ، بين المقادير N_i ، العلاقة الإضافية التالية :

$$\frac{1}{N} \sum_{i=1}^{k} N_i x_i = np .$$

كذلك بالنسبة لتسوية قانون بواسّون (Poisson) الذي نقدّر متغيّره الوسيطي m بواسطة المعدّل الوسطي الملحوظ $\bar{x}$ ، لدينا العلاقة التالية :

$$\frac{1}{N}\sum_{i=1}^{k} N_i x_i = m .$$

في هاتين الحالتين ، r تساوي 1 ويكون عدد درجات الحرّية بالتالي :

$$\nu = k - 2 .$$

بالنسبة لتسوية القانون الطبيعي الذي نقدّر متغيّره الوسيطي m بواسطة المعدّل الوسطي الملحوظ $\overline{x}$ وانحرافه النموذجي σ بواسطة الانحراف النموذجي الملحوظ s ، يوجد بين المقادير N_i العلاقتان الإضافيتان التاليتان :

$$\frac{1}{N}\sum_{i=1}^{k} N_i x_i = m ,$$

$$\frac{1}{N}\sum_{i=1}^{k} N_i (x_i - m)^2 = \sigma^2 .$$

في هذه الحالة ، r تساوي 2 ، ويكون عدد درجات الحرية : $\nu = k-3$.

يستند اختبار χ^2 ، الذي أدرجه ك. بيرسون K. Pearson ، إلى طريقة التفكير التالية[1] :

نضع الفرضية التي تقول بأنّ الظاهرة الملحوظة تتبع القانون النظري المفترض P ؛ في هذه الشروط ، تكون المسافة D ، بحكم التقلّبات العشوائية ، متغيّرة χ^2 ذات $\nu = k - r - 1$ درجة حريّة .

- إذا كان هناك احتمال قوي (نجده عن طريق جدول χ^2 لأن تأخذ D قيمة أكبر من القيمة d الملحوظة ، عندها تكفي التقلّبات العشوائية لتفسير المسافة المسجّلة : ونحكم على الفرضية بالقبول[2] .

- أمّا إذا لم يكن هناك سوى احتمال ضعيف (مثلاً ، احتمال أصغر من 5%) للحصول على قيمة D أكبر من القيمة d الملحوظة ، من الممكن جدّاً أن تكون هذه القيمة المرتفعة عائدة إلى عدم موافقة القانون النظري P : عندها نرمي الفرضية التي تعتبر أن الظاهرة الملحوظة تتبع هذا القانون .

(1) إنّ طريقة التفكير الإحصائية التي تحمل اسم « اختبار الفرضيات » موسّعة في الفصل VI ، القسم II

(2) وهذا لا يعني أنّ الفرضية صحيحة بالضرورة ، ولكن فقط أنّ للمعلومات التي بحوزتنا لا تسمح لنا برميها . ونشير إلى أنّه من الممكن أن نحكم ، من هذا المنظار ، بالقبول على عدّة قوانين نظرية لتمثيل نفس مجموعة الحالات .

ملاحظات عملية

1 ـ كي تتبع المسافة D القانون χ^2 ، تستلزم شروط الميل أن لا تكون المقادير النظرية Npi لمختلف الفئات صغيرة جدّاً : عملياً ، نعتبر أنّها يجب أن تكون على الأقلّ مساوية لـ 4 أو 5 .

بالتالي ، يكون بعض الأحيان من الضروري تجميع بعض الفئات ، بصورة خاصّة عند طرفي التوزيع . وطبعاً يكون عدد الفئات k الذي يجب أخذه بعين الاعتبار عند حساب عدد درجات الحرّية هو عدد الفئات بعد التجميع .

2 ـ عادةً ، تكون درجات الاحتمال التي نقرّر بعدها أن نرمي الفرضية بين 2 و5% .

3 . أمثلة

المثل 1 . القانون ذو الحدّين

لنعد إلى مثل تسوية قانون ذي حدّين مع توزيع 100 عيّنة حسب عدد القطع المعيبة (الفصل II ، القسم I ، ص 78) .

1 . لنفترض أنّه لدينا مسبقاً أسباب تجعلنا نفكّر أنّ نسبة القطع المرفوضة المئوية في الكمّية المصنوعة هي 4% (مثلاً ، حسب إرشادات صانع الآلة) . إذن سنختبر الفرضية التي تقول أن توزيع N=100 كمّية من n=40 قطعة يتبع قانوناً ذا حدّين بمتغيّرين وسيطين n = 40 و $p = 0{,}04$ ؛ $\mathcal{B}(40 ; 0{,}04)$.

يتمّ حساب المسافة d بين التوزيعين التجريبي والنظري على الجدول 10 . وقد قمنا بتجميع الكيفية الأخيرة ، ذات مقدار نظري أصغر من 4 ، مع سابقتها . يتمّ الحساب إذن على 5 كيفيات فقط ، فنحصل على :

d = 11,62 مع $\nu = 5 - 1 = 4$ درجة حريّة ، وذلك لأنّه لم يتمّ تقدير أيّ متغيّر وسيطي إنطلاقاً من الحالات الملحوظة . غير أنّ جدول χ^2 يعطينا (الملحق ، الجدول 4) .

$$P\{\chi^2 \geqslant 11{,}67\} = 0{,}02 .$$

ليس لدينا إذن سوى فرصتين على 100 تقريباً أن نتجاوز قيمة d المحسوبة بفعل مجرّد التقلّبات العشوائية . بما أنّ هذا الاحتمال ضعيف ، نرمي فرضية القانون ذي الحدّين $\mathcal{B}(40 ; 0{,}04)$.

2 . لنعد الآن إلى طريقة التسوية العادية ونأخذ القانون ذا الحدّين الذي يساوي أمله الرياضي معدّل التوزيع الملحوظ الوسطي : $\bar{x} = 1{,}2$. سوف نختبر الفرضية

التي تقول أنّ توزيع الكمّيات الـ 100 يتبع هذا القانون ذا الحدّين بمتغيرين وسيطين n=40 وp=0,03 :

يتمّ حساب المسافة d في الجدول 11 . اضطررنا هذه المرّة إلى تجميع الكيفيات الأخيرة الثلاث في فئة واحدة كي يصبح مقدارها النظري كافياً . إذن يتمّ الحساب على أربع كيفيّات لنحصل على :

d = 0,50 ، مع $\nu = 4 - 1 - 1 = 2$ درجتي حرّية ، وذلك لأنّنا قمنا بتقدير المتغيّر الوسيطي p انطلاقاً من الحالات الملحوظة .

يعطينا الجدول χ^2 : $P\{\chi^2 \geq 0,45\} = 0,80$ ،

الجدول 10 . اختبار تسوية قانون ذي حدّين $\mathcal{B}(40;0,04)$

عدد القطع المعيبة x_i	المقادير الملحوظة N_i	المقادير النظرية Np_i	$N_i - Np_i$	$(N_i - Np_i)^2$	$\frac{(N_i - Np_i)^2}{Np_i}$
0	28	19,5	8,5	72,25	3,71
1	40	32,6	7,4	54,76	1,68
2	21	26,4	− 5,4	29,16	1,10
3	7	14,0	− 7,0	49,00	3,50
4	3 } 4	5,4 } 7,5	− 3,5	12,25	1,63
5 وأكثر	1	2,1			
المجموع	100	100,0	—	—	d = 11,62

الجدول 11 . اختبار تسوية قانون ذي حدّين $\mathcal{B}(40; 0,03)$

(القراءة من اليسار إلى اليمين)

عدد القطع المعيبه x_i	المقادير الملحوظة N_i	المقادير النظرية Np_i	$N_i - Np_i$	$(N_i - Np_i)^2$	$\frac{(N_i - Np_i)^2}{Np_i}$
0	28	29,6	− 1,6	2,56	0,09
1	40	36,6	+ 3,4	11,56	0,32
2	21	22,1	− 1,1	1,21	0,05
3	7	8,6			
4	3 } 11	2,5 } 11,7	− 0,7	0,49	0,04
5 وأكثر	1	0,6			
المجموع	100	100,0	—	—	d = 0,50

لـدينا 80 فـرصة عـلى 100 أن نتجاوز القيمـة المحـسوبـة بفعـل مجـرّد التقلّبـات العشوائية . إذن الفرضية التي تقول أن الحالات الملحوظة تتبع قانوناً ذا حدّين بمتغيّرين وسيطيّين n=40 وp=0,03 هي فرضية مقبولة . بعبارة أخرى ، آخـذين بعين الاعتبـار المعلومات التي بحوزتنا ، لا شيء يسمح لنا بدحض هذه الفرضية .

ملاحظة . في هذا المثل ، يجب الانتبـاه من الخلط بين n ، وهي مقـدار كلّ من الكمّيـات وN ، وهي عدد الكمّيـات موضـع الدراسـة . يتبع عـدد القـطع التي يجب رفضها ، في كلّ كمّية ، قانـوناً ذا حـدّين $\mathscr{B}$ (n,p) ، ويسمح هـذا القانـون بحساب الاحتمال النظري p_i لأن نجد x_i قطعة معيبة . يجب أن نأخذ N كمقدار عيّنة الكمّيات (هنا N=100) ، المسحوبة من المجتمع الإحصائي النظري اللامتناهي للكمّيات التي تتبع القانون $\mathscr{B}$(n,p) . إذن المقدار النظري الذي يطابق x_i قطعة معيبة يساوي Np_i .

المثل 2 . قانون بواسّون

لنختبر على نفس المثل تسوية قانون بواسّون يساوي أمله الرياضي المعدّل الوسطي للتوزيع الملحوظ (أنظر الفصل II ، القسم III ، ص 98) : $\mathscr{P}$(1,2) .

يتمّ حساب المسافة d على الجدول 12 ، وقد جمّعنا الكيفيات الثلاث الأخيرة في فئة واحـدة كي يصبـح مقـدارهـا كبيـراً بمـا فيـه الكفـايـة ، نحصل على :

d=0,69 ، مع $\nu = 4 - 1 - 1 = 2$ درجتي حرّية ، وذلك لأنّنا قمنا بتقدير متغيّر قانون بواسّون الوسيطي m انطلاقاً من الحالات الملحوظة .

يعطينا جدول قانون χ^2 (الملحق ، الجدول 5) : $P(\chi^2 \geqslant 0,71) = 0,70$.

عدد القطع المعيبة x_i	المقادير الملحوظة N_i		المقادير النظرية Np_i		$N_i - Np_i$	$(N_i - Np_i)^2$	$\frac{(N_i - Np_i)^2}{Np_i}$
0	28		30,1		− 2,1	4,41	0.15
1	40		36,1		+ 3,9	15,21	0.42
2	21		21,7		− 0,7	0,49	0.02
3	7	11	8,7	12,1	− 1,1	1,21	0,10
4	3		2,6				
5 وأكثر	1		0,8				
المجموع	100		100,0		—	—	d=0,69

الجدول 12 . اختبار تسوية قانون بواسّون $\mathscr{P}$ (1, 2)

لدينا 70 فرصة على 100 أن تتجاوز ، بفعل التقلّبات العشوائية فقط ، القيمة المحسوبة : إذن فرضية قانون بواسّون ذي متغيّر وسيطي m=1,2 هي فرضية مقبولة .

نلاحظ أنّنا حكمنا بالقبول على قانونين مختلفين ، القانون ذي الحدّين (40; 0,03)ℬ وقانون بواسّون (1,2)𝒫 ، لتمثيل الظاهرة . لا عجب في هذه الحالة الخاصّة لأنّ قانون بواسّون يبدو فيها وكأنّه تقريب للقانون ذي الحدّين . ولكن بشكل عام ، قد نعتبر عدة تسويات ذات طبيعة مختلفة صالحة ، من وجهة نظر الاختبار ، لتمثيل نفس مجموعة الحالات الملحوظة : لا يجب أن ننسى أن فرضية مقبولة ليست بالضرورة فرضية صحيحة .

المثل 3 . القانون الطبيعي

لننتقل إلى مثل تسوية قانون طبيعي مع التوزيع الملحوظ لأقطار 400 برغي (القسم I ، ص 126) . سوف نختبر صحة تسوية القانون الطبيعي ذي المتغيّرين الوسيطيّين m=3,32 و0,10= σ اللذين يطابقان على التوالي معدّل التوزيع الملحوظ الوسطي وانحرافه النموذجي : N (3,32; 0,10) .

يتمّ حساب المسافة d بين التوزيعين التجريبي والنظري على الجدول 13 . وقد قمنا بتجميع الفئتين الأوليّين والفئتين الأخيرتين بشكل لا يعود معه المقدار النظري .

الجدول 13 . اختبار تسوية القانون الطبيعي N (3,32;0,10) (القراءة من اليسار إلى اليمين) .

فئات الأقطار $(e_{i-1}-e_i)$	المقادير الملحوظة N_i		المقادير النظرية Np_i		N_i-Np_i	$(N_i-Np_i)^2$	$\frac{(N_i-Np_i)^2}{Np_i}$
3,00-3.05	3	9	1.4	5.6	3.4	11.56	2,06
3.05-3.10	6		4.2				
3,10-3.15	13		12.2		0.8	0.64	0.05
3,15-3.20	23		28.2		− 5.2	27.04	0.96
3.20-3.25	39		50.8		− 11.8	139.24	2.74
3.25-3.30	78		71.5		6.5	42.25	0.59
3.30-3.35	91		78.9		12.1	146.41	1.86
3.35-3.40	72		68.0		4.0	16.00	0.24
3.40-3.45	42		46.1		− 4.1	16.81	0.36
3.45-3.50	17		24.3		− 7.3	53.29	2.19
3.50-3.55	9		10.1		− 1.1	1.21	0.12
3.55-3.60	5	7	3.3	4.3	2.7	7.29	1.70
3.60-3.65	2		1.0				
المجموع	100		100.0				$d=12.87$

لا يعود معه المقدار النظري لأي فئة أصغر من 4 ، فنجد :

$d = 12,87$ ، مع $\nu = 11-2-1=8$ درجات حرّية ، وذلك لأنّنا قمنا بتقدير المتغيّرين الوسيطيّين m و انطلاقاً من الحالات الملحوظة .

يعطينا جدول χ^2 (الملحق ، الجدول 5) .

$$P\{\chi^2 \geqslant 11,03\} = 0,20$$
$$P\{\chi^2 \geqslant 13,36\} = 0,10 .$$

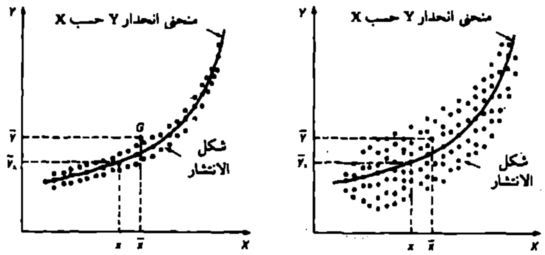
إذا كانت الظاهرة الملحوظة تخضع فعلاً للقانون الطبيعي $N(3,32;0,10)$ لدينا إذن أكثر من 10 فرص على 100 كي نتجاوز قيمة d المحسوبة ، بفعل التقلّبات العشوائية فقط . وهذا الاحتمال هو أكبر من أن نسمح لأنفسنا برمي الفرضية : نعتبرها إذن مقبولة ، معتمدين على المعلومات المتوفّرة لدينا . غير أنّه احتمال ضعيف كي يكون لقبول الفرضية معنوية كبيرة .

الفصل الرابع

الانحدار والارتباط

لقد عرضنا بعض مبادىء تحليل السلاسل الإحصائية ذات البعدين في كتاب « الإحصاء الوصفي » (الفصل III) . بصورة خاصّة ، يسمح لنا حساب التوزيعات الهامشية والشرطية بتحويل توزيع ذي بعدين إلى مجموعة توزيعات ذات بعد واحد يمكننا تمثيلها بيانياً ونلخّصها عددياً بواسطة مقاييسها ذات النزعة المركزية ومقاييس التشتّت .

لكنّ هذا الأمر لا يمكّننا من الإلمام بشكل كاف بمجمل الفكرة الموجودة في توزيع متغيّرتين . بالفعل ، إنّ تمثيل هذه التوزيعات البياني ، يبرز فكرة جديدة هي فكرة التبعيّة الإحصائية أو الارتباط بين متغيّرتين ملحوظتين : عندما تكون المتغيّرة Y مرتبطة بالمتغيّرة X ، فإنّ النقاط التي تمثّل أزواج القيم (x_i, y_i) تؤلّف شكل انتشار متراوح الطول والامتداد (الشكل 37) . فمعرفة قيمة تأخذها X تحمل لنا فكرة إضافية حول القيم التي تأخذها Y : إذا كانت X تساوي x ، فإنّ Y تأخذ بالمتوسّط القيمة $\bar{y}_x$ وليس $\bar{y}$.



الشكل 37 . الارتباط بين متغيّرتين : شكل وكثافة العلاقة

عندما تكون المتغيّرة Y مرتبطة بالمتغيّرة X ، تنطرح لدينا مشكلتان :

ـ تحديد شكل العلاقة الإحصائية الموجودة بين Y وX : أي تحديد منحنى انحدار Y تبعاً لـ X .

ـ قياس كثافة العلاقة بواسطة مُعامِل ملائم . في الواقع ، إذا قمنا بمقارنة الرسمين البيانيّين على الشكل 37 ، نستنتج أن منحني انحدار Y حسب X متشابهان في الرسمين . إلّا أنّ كثافة العلاقة تبدو بوضوح مرتفعة في الأوّل أكثر من الثاني . المعامل الذي يمكّننا من قياس درجة العلاقة هذه هو ، تبعاً للحالة ، نسبة الارتباط أو معامل الارتباط الخطّي .

القسم I

المقاييس الهامشية والشرطية لتوزيع متغيّرتين

1 . المقاييس الهامشية .ـ 2 . المقاييس الشرطية .ـ 3 . التغاير . ـ 4 .ـ العلاقات بين المقاييس الهامشية والشرطية .

لنأخذ توزيع المجتمع الإحصائي P ، وحجمه الكلّي n ، حسب المتغيّرتين الإحصائيتين X وY (الجدول 14) . لقد حدّدنا التوزيعات الهامشية والشرطية للمتغيّرتين X وY في كتاب « الإحصاء الوصفي » ، الفصل الثالث. من الطبيعي أن يخطر لنا حساب مقاييس النزعة المركزية ومقاييس التشتّت لهذه التوزيعات ذات البعد الواحد .

1 . المقاييس الهامشية

إنّ العامود الهامشي في الجدول 14 والذي يتضمّن المقادير $n_{i.}$ التي تطابق كل قيمة x_i تأخذها المتغيّرة X ، هو توزيع X الهامشي . ومقياسا X الهامشيان (المتوسّط والتباين) هما :

$$\bar{x} = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.}\, x_i$$

$$V(X) = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2 .$$

كذلك ، فإنّ السطر الأخير من الجدول 14 ، والذي يتضمّن المقادير $n_{.j}$ ، هو

توزيع Y الهامشي . بالتالي :

		P_1	P_2	$\ldots$	P_j	$\ldots$	P_l	
المجتمعات الإحصائية الثانوية ↓	المتغيّرة Y / المتغيّرة X	y_1	y_2	$\ldots$	y_j	$\ldots$	y_l	حواصل الجمع
P'_1	x_1	n_{11}	n_{12}	$\ldots$	n_{1j}	$\ldots$	n_{1l}	$n_{1.}$
P'_2	x_2	n_{21}	n_{22}	$\ldots$	n_{2j}	$\ldots$	n_{2l}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
P'_i	x_i	n_{i1}	n_{i2}	$\ldots$	n_{ij}	$\ldots$	n_{il}	$n_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
P'_k	x_k	n_{k1}	n_{k2}	$\ldots$	n_{kj}	$\ldots$	n_{kl}	$n_{k.}$
	حواصل الجمع	$n_{.1}$	$n_{.2}$	$\ldots$	$n_{.j}$	$\ldots$	$n_{.l}$	$n_{..}$

الجدول 14 . التمثيل العام لتوزيع احصائي بمتغيّرتين

إذا كانت X (أو Y) متغيّرة متواصلة فإنّنا نختار x_i (أو y_j) مساويةً لمركز الفئة المناسبة ، كما بالنسبة لحساب متوسّط السلاسل الإحصائية ذات المتغيّرة الواحدة وانحرافها النموذجي .

$$\bar{y} = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j}\, y_j$$

$$V(Y) = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2 .$$

مثلاً . فيما يلي ، سنأخذ كمثل توزيع عمّال مصانع شركة معيّنة حسب العمر والراتب الشهري ، وقد تمّ عرض هذا المثل في كتاب « الإحصاء الوصفي » ، الفصل الثالث ، سنسترجعه هنا في الجدول 15 .

يعطينا العامود (1) (في الجدول 15 توزيع العمّال الهامشي حسب الراتب الشهري R . ونتيجة حساب مقياسي هذا التوزيع هي التالية[1] :

ـ متوسّط R الهامشي : $\bar{r} = \frac{1}{n} \sum_{i=1}^{k} n_{i.} r_i = 1\,008{,}6$ F.

(1) نوصي القارىء بأن يقوم بنفسه بهذه الحسابات (التي لم نفصّلها هنا) حسب الطرق المعروضة في كتاب « الإحصاء الوصفي » ، الفصل V .

ـ تباين R الهامشي :

$$V(R) = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(r_i - \bar{r})^2 = 36\ 350$$

$$\sigma_R = 190{,}7\ F.$$

يعطينا السطر (2) من هذا الجدول توزيع نفس هؤلاء العمّال الهامشي حسب العمر A . قيمة مقياسي هذا التوزيع هي[1] :

ـ متوسّط A الهامشي :

37,4 سنة

$$\bar{a} = \frac{1}{n}\sum_{j=1}^{l} n_{.j}\, a_j = 37{,}4$$

ـ تباين A الهامشي :

$$V(A) = \frac{1}{n}\sum_{j=1}^{l} n_{.j}(a_j - \bar{a})^2 = 84{,}22$$

9,2 سنة

$$\sigma_A = 9{,}2$$

2 . المقاييس الشرطية

إنّ العامود j من الجدول 14 ، والذي يصف توزيع n_{ij} وحدةً إحصائية تمثّل القيمة y_j التي تأخذها المتغيرة Y وذلك حسب المتغيّرة X ، هو توزيع X الشرطي المتعلّق بـ $Y=y_j$. مقياسا (متوسّط وتباين) المتغيّرة الشرطية X/y_j هما إذاً :

$$\bar{x}_j = \frac{1}{n_{.j}}\sum_{i=1}^{k} n_{ij}\, x_i$$

$$V_j(X) = \frac{1}{n_{.j}}\sum_{i=1}^{k} n_{ij}(x_i - \bar{x}_j)^2$$

كذلك ، فإنّ السطر i من الجدول 14 يصف توزيع n_{ij} وحدة إحصائية تمثّل القيمة x_i التي تأخذها المتغيّرة X وذلك حسب المتغيّرة Y ، وهو عبارة عن توزيع Y الشرطي المتعلّق بـ $X=x_i$. المقياسان الشرطيان المناسبان هما إذاً :

$$\bar{y}_i = \frac{1}{n_{i.}}\sum_{j=1}^{l} n_{ij}\, y_j$$

$$V_i(Y) = \frac{1}{n_{i.}}\sum_{j=1}^{l} n_{ij}(y_j - \bar{y}_i)^2$$

(1) انظر الملاحظة السابقة .

الجدول 15 . عدد العمّال موزّعين حسب العمر والراتب الشهري . كانون الثاني (يناير) 1970

المصدر : دائرة الموظّفين

العمر / الراتب الشهري	أقلّ من 25 سنة	من 25 إلى أقلّ من 30 سنة	من 30 إلى أقلّ من 35 سنة	من 35 إلى أقلّ من 40 سنة	من 40 إلى أقلّ من 45 سنة	م 45 إلى أقلّ من 50 سنة	من 50 إلى أقلّ من 55 سنة	55 سنة وأكثر	حواصل الجمع (1)
أقلّ من 800 F	207	121	38	17	10	2	7	3	405
من 800 إلى أقلّ من 900F	302	461	513	103	86	6	10	2	1 483
من 900 إلى أقل من 1 000F	18	526	682	567	613	431	105	60	3 002
من 1000 إلى أقلّ من 1 200F		111	342	298	416	480	226	37	1 910
من 1200 إلى أقلّ من 1 500F		1	3	182	227	263	98	18	792
من 1 500 إلى أقلّ من 2 000F				18	22	13	12	5	70
2000F وأكثر				1	14	6	7	5	33
حواصل الجمع (2)	527	1220	1578	1186	1388	1201	465	130	7695

(1)S يبيّن هذا العامود توزيع العمّال الماشي حسب الراتب الشهري .
(2) يبيّن هذا السطر توزيع العمّال الماشي حسب العمر .

مثلاً . تعطينا عواميد الجدول 15 توزيعات العمّال الشرطية حسب الراتب الشهري متعلّقاً بالعمر . لكلّ من هذه العواميد، يمكننا حساب متوسّط وتباين الراتب . يعطينا الجدول 16 قيم هذه المقاييس .

الجدول 16 . مقاييس الراتب الشرطية تبعاً للعمر

فئة العمر	مركز الفئة	الراتب الشهري المتوسّط (بالفرنك)	تباين الراتب الشهري وانحرافه النموذجي	
	a_j	$\overline{r_j}$	$V_j(R)$	$\sigma_j(R)$
25 سنة	20,0	794,5	6100	78,1
30 سنة	27,5	901,5	9825	99,1
35 سنة	32,5	944,5	9875	99,4
40 سنة	37,5	1050,0	33000	181,7
45 سنة	42,5	1077,5	43575	208,8
50 سنة	47,5	1111,5	32825	181,2
55 سنة	52,5	1141,0	49450	222,4
	60,0	1119,5	86100	293,4
المقاييس الهامشية (لكلّ الأعمار)		$\overline{r} = 1\,008,5$	$V(R) = 36\,347$	$\sigma_R = 190,7$

نلاحظ مثلاً أنّ الانحرافات النموذجية بالنسبة للموظّفين الشبّان هي أضعف منها بالنسبة للموظّفين الأكبر سنّاً : إذ من الطبيعي أن يكون المجتمع الإحصائي الشابّ متجانساً أكثر من ناحية الرواتب .

بالمقابل ، تعطينا أسطر الجدول 15 توزيعات الموظّفين الشرطية حسب العمر متعلّقاً بالراتب الشهري . ويعرض الجدول 17 المقاييس الشرطية المناسبة .

ملاحظة : في هذه الحسابات اعتبرنا كلّ المشاهدات مجمّعة في مراكز الفئات المختلفة . وقد تمّ تحديد « مركز » الفئتين الطرفين اصطلاحياً بقيمة قريبة من متوسّط الفئة المفترض .

الجدول 17 . مقاييس العمر الشرطية تبعاً للراتب

فئة الراتب الشهري (بالفرنك)	مركز الفئة r_i	متوسّط العمر $\bar{a}_i$	تباين العمر $V_i(A)$	والانحراف النموذجي $\sigma_i(A)$
800	700	25,7	58,23	7,6
900	850	29,6	43,11	6,6
1000	950	37,1	62,30	7,9
1200	1100	41,8	58,50	7,7
1500	1350	44,6	29,68	5,5
2000	1750	45,1	43,31	6,6
	2200	48,0	42,34	6,5
المقاييس الهامشية (لكلّ الرواتب)		$\bar{a} = 37,4$	V(A) = 84,32	$\sigma A = 9,2$

3 . التغاير

قياساً على المتغيّرات العشوائية (أنظر الفصل I ، القسم V ، ص 62) ، نحدّد تغاير (covariance) متغيّرتين احصائيتين X وY بواسطة :

$$\text{cov}\,(XY) = \frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})\,(y_j - \bar{y})\,.$$

نلاحظ أن هذا التحديد يتلاحم مع تحديد التباين : إذا جعلنا y=x ، نحصل مجدّداً على قاعدة التباين .

يساوي التباين صفراً إذا كانت المتغيّرتين مستقلّتين . سوف تدخل هذه الكمّية في دراسة العلاقة بين متغيّرتين ولا سيّما في دراسة الارتباط الخطّي .

الحساب العملي

لتسهيل حساب التغاير ، نستعمل طرقاً شبيهة بالطرق المستعملة في حساب التباين : القاعدة المتبسّطة واستبدالات المتغيّرة (أنظر كتاب « الإحصاء الوصفي » ، الفصل V) .

ـ القاعدة المتبسّطة

من الممكن بسط (توسيع) قاعدة التحديد للحصول على عبارة متكيّفة أكثر مع

الحساب العددي :

$$\operatorname{cov}(XY)=\frac{1}{n}\sum_i\sum_j n_{ij}(x_i-\bar{x})(y_j-\bar{y})$$

$$=\frac{1}{n}\sum_i\sum_j n_{ij}x_iy_j-\bar{y}.\frac{1}{n}\sum_i\sum_j n_{ij}x_i-\bar{x}.\frac{1}{n}\sum_i\sum_j n_{ij}y_j+\overline{xy}\frac{1}{n}\sum_i\sum_j n_{ij}$$

إلاّ أنّه ، إنطلاقاً من التعريف :

$$\sum_i\sum_j n_{ij}=n$$

$$\sum_i\sum_j n_{ij}x_i=\sum_i n_{i.}x_i=n\bar{x}$$

$$\sum_i\sum_j n_{ij}y_j=\sum_j n_{.j}y_j=n\bar{y}$$

إذاً :

$$\operatorname{cov}(XY)=\frac{1}{n}\sum_i\sum_j n_{ij}x_iy_j-\overline{xy}.$$

ـ استبدال المتغيّرة

إنّ استبدال متغيّرة ملائم نجريه على x وy يعطينا غالباً تسهيلاً إضافياً للحسابات . لنختر بالنسبة لِـ x وy نقطتي أصل جديدتين x_0 وy_0 ووحدتي قياس جديدتين α وβ ، بشكل تكون فيه المتغيّرتان المساعدتان x′ وy′ عددين صحيحين أبسط من x وy :

$$x'_i=\frac{x_i-x_0}{\alpha},\qquad y'_j=\frac{y_j-y_0}{\beta}$$

أي :

$$x_i=\alpha x'_i+x_0,\qquad y_j=\beta y'_j+y_0.$$

بفضل خصائص المتوسّط الحسابي ، نجد نفس العلاقتين بين المتوسّطات :

$$\bar{x}=\alpha\bar{x}'+x_0,\qquad \bar{y}=\beta\bar{y}'+y_0.$$

وبالطرح :

$$x_i-\bar{x}=\alpha(x'_i-\bar{x}'),\qquad y_j-\bar{y}=\beta(y'_j-\bar{y}').$$

إذاً :

$$\operatorname{cov}(XY) = \frac{1}{n} \sum_i \sum_j n_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \alpha\beta \frac{1}{n} \sum_i \sum_j n_{ij}(x'_i - \bar{x}')(y'_j - \bar{y}')$$
$$= \alpha\beta \operatorname{cov}(X' Y').$$

سنجد لاحقاً ، حول موضوع التسوية الخطّية ، أمثلة عن حساب التغاير (القسم III ، ص 190 و196) . .

4 . العلاقات بين المقاييس الهامشية والشرطية

يمكننا اعتبار المجتمع الإحصائي P مؤلّفاً :

- إمّا من l مجتمعاً ثانوياً $P_1, P_2, \ldots, P_l$ بمقادير $n_{.1}$ ، $n_{.2}$ ، . . . ، $n_{.l}$ ، تناسب توزيعات X الشرطية متعلّقة بـ Y ؛
- إمّا من k مجتمعاً ثانوياً ت$P_1, P_2, \ldots, P_k$ بمقادير $n_{1.}$ ، $n_{2.}$ ، . . . ، $n_{k.}$ ، تناسب توزيعات Y الشرطية متعلّقة بـ X .

يمكننا إذن أن نطبّق على متوسّط وتباين X أو Y ، الهامشي النتائج التي بيّناها في كتاب « الإحصاء الوصفي » ، الفصل V ، والتي تتعلّق بعبارة متوسّط وتباين مجتمع إحصائي يتألّف من علّة مجتمعات ثانوية .

عبارة المتوسّط الهامشي تبعاً للمتوسّطات الشرطية

إنّ متوسّط مجمل المجتمع الإحصائي يساوي متوسّط متوسّطات المجتمعات الثانوية مرجّحاً (نتيجة من كتاب « الإحصاء الوصفي » ، الفصل V ، القسم I ، الفقرة 3.C) . ومعاملات الترجيح هي نسب المجتمعات الثانوية في المجتمع الكلّي .

إذا اعتبرنا توزيع X الهامشي مؤلّفاً من توزيعات X الشرطية متعلّقة بـ Y ، نحصل على عبارة متوسّط X الهامشي تبعاً لمتوسّطات X الشرطية متعلّقة بـ Y :

$$\bar{x} = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j}\, \bar{x}_j .$$

كذلك ، إذا اعتبرنا توزيع Y الهامشي مؤلّفاً من توزيعات Y الشرطية متعلّقة بـ X ، نحصل على عبارة متوسّط Y الهامشي تبعاً لمتوسّطات Y الهامشية متعلّقة بـ :

$$\bar{y} = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.}\, \bar{y}_i .$$

المتوسّـط الهامشي يسـاوي المتوسّـط المرجّـح للمتوسّـطات الشرطية .

عبارة التباين الهامشي تبعاً للمتوسّطات والتباينات الشرطية

إنّ تباين مجمل المجتمع الإحصائي يسـاوي حاصـل جمع عنصـرين ، المتوسّـط المرجّح لتباينات المجتمعات الثانوية والتباين المـرجّـح لمتوسّطات المجتمعـات الثانـوية (نتيجة من « الإحصاء الوصفي » ، الفصـل V ، القسم II ، الفقرة 4.C) .

إذاً ، إذا أخذنا توزيع X الهامشي مؤلّـفاً من توزيعات X الشـرطية متعلّـقة بـ Y ، نحصل على :

$$V(X) = \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j}\, V_j(X) + \frac{1}{n_{..}} \sum_{j=1}^{l} n_{.j}(\bar{x}_j - \bar{x})^2 .$$

كذلك ، إذا أخذنا توزيع Y الهامشي مؤلّـفاً من توزيعات Y الشرطية متعلّـقة بـ X :

$$V(Y) = \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.}\, V_i(Y) + \frac{1}{n_{..}} \sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2 .$$

التباين الهامشي يساوي حاصل جمع متوسّـط التباينات الشرطية المرجّـح مع تباين توسّـطات الشرطية المرجّـح .

إذاً ، ينتج تشتّت التوزيع الهامشي عن عاملين :

- تشتّت كلّ من التوزيعات الشرطية حول متوسّـطها ؛
- تشتّت المتوسّـطات الشرطية فيما بينها .

هكذا يمكننا تفسير قسـم من تباين X (أو Y) الكلّي بتباين المتوسّـطات الشرطية (العنصر الثـاني) ، أمّـا التباين المتوسّـط الناتج عن التنافرات الخـاصّـة بكـلّ من التوزيعات الشرطية (العنصر الأوّل) فيبدو كتباين متبقّ . على أساس هذه التجزئة للتباين الكلّي سنبني تعريف نسبة الارتباط .

سنرى مثلاً عن تجزئة التباين الهامشي في إطار حساب نسبة الارتباط (القسم II ، ص 184) .

القسم II

منحنيات الانحدار ونسبة الارتباط

1 . منحنيات الانحدار : A . تعريف ؛ B . المعنى .ـ 2 . نسبة الارتباط : A . تعريف ؛ B . الخصائص ؛ C . الحساب العملي .ـ 3 . مبدأ طريقة المربّعات الصغرى .

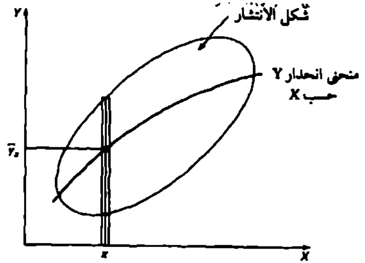
يبرز لنا تمثيل توزيع العمّال حسب العمر والراتب الشهري بيانياً (أنظر كتاب « الإحصاء الوصفي » ، الفصل III ، الشكل 25) وجود علاقة إحصائية بين هاتين المتغيّرتين . ويهدف تحديد منحنى الإنحدار الى تعيين شكل هذه العلاقة ، فيما يسمح لنا حساب نسبة الارتباط بقياس كثافتها .

1 . منحنيات الانحدار

A . تعريف

لنعد إلى الحالة العامّة حيث توزيع مجتمع إحصائي P حسب المتغيرتين X وY .

تتركّب العلاقة الإحصائية التي تربط المتغيّرة Y بالمتغيّرة X بواسطة منحنى تغيّر المتوسّطات الشرطية $\bar{y}_x$ تبعاً للقيم x التي تأخذها متغيّرة العلاقة . نسمّي هذا المنحنى منحنى انحدار Y حسب X (الشكل 38) .



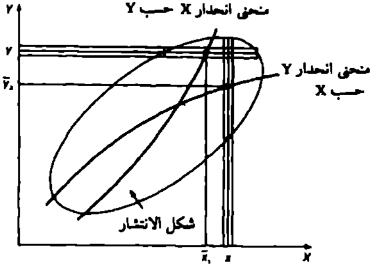
الشكل 38 . منحنى انحدار Y حسب X

وبالمقابل ، منحنى انحدار X حسب Y هو منحنى تغيّر المتوسّطات الشرطية $\bar{x}_y$.

تبعاً للقيم y التي تأخذها متغيّرة العلاقة ، وهو يعبّر عن العلاقة التي تربط المتغيّرة X بالمتغيّرة Y . يميّز إذاً التوزيع بمتغيّرتين بواسطة منحنيي انحدار (الشكل 39) .

لنأخذ منحنى انحدار Y حسب X .

المتغيّرة المنفصلة : إذا كانت متغيّرة العلاقة X منفصلة فإنّ منحنى انحدار Y حسب X ، في الحقيقة ، يتألّف من متالية النقاط التي تناسب المتوسّطات الشرطية $\bar{y}_i$ المتعلّقة بالقيم x_i المنفصلة التي تأخذها متغيّرة العلاقة .



الشكل 39 . منحنيا انحدار توزيع بمتغيّرتين

المتغيّرة المتواصلة : إذا كانت متغيّرة العلاقة متواصلة فإنّ منحنى الانحدار هو منحنى حقيقي . إلاّ أنّه على الصعيد العملي تتجمّع المشاهدات ضمن فئات . اصطلاحياً ، ننسب المتوسّطات الشرطية $\bar{y}_i$ التي توافق مختلف فئات متغيّرة العلاقة X ، إلى مركز الفئة المناسبة x_i . إذاً ، لا نحيط علماً ، في الحقيقة ، إلاّ ببعض نقاط منحنى الإنحدار ، وهي النقاط التي تطابق مراكز فئات متغيّرة العلاقة .

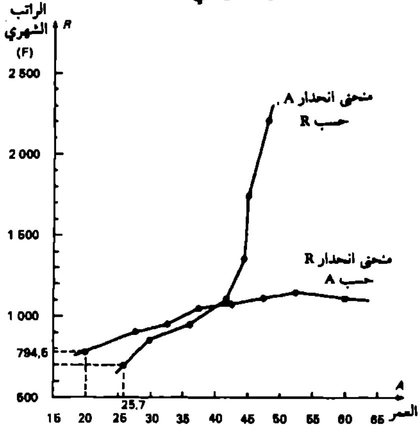
مثلاً . يسمح لنا الجدولان 16 و17 برسم منحنيي انحدار توزيع الموظّفين حسب الراتب الشهري والعمر (الشكل 40) .

نرسم منحنى انحدار الراتب R حسب العمر A انطلاقاً من النقاط التي تطابق مراكز مختلف فئات العمر a_i على المحور السيني ، والرواتب المتوسّطة المناسبة $\bar{r}_i$ على المحور الصادي .

ونرسم منحنى انحدار A حسب R انطلاقاً من النقاط التي تطابق مراكز مختلف

فئات الرواتب n على المحور الصادي ، والأعمار المتوسّطة المناسبة $\bar{a}_i$ على المحور السيني .

إن منحنيات الانحدار لا تلخّص كلّ المعلومات التي يحتويها توزيع متغيرتين . ففي الواقع ، يتميّز كلّ من التوزيعات الشرطية ليس فقط بقيمته المركزية (المتوسّط الشرطي) ؛ بل أيضاً بتشتّته (التباين الشرطي) . لا يتضمّن منحنى الإنحدار ، الذي يمثّل تغيّر المتوسّطات الشرطية ، أي فكرة عن التشتّتات . نسبة الارتباط هي ما سيعطينا قياساً لتشتت التوزيعات الشرطية الوسطي .



الشكل 40 . منحنيا انحدار توزيع الموظّفين حسب العمر A والراتب الشهري R

B . معنى منحنيات الانحدار

يدخل منحنيا انحدار توزيع متغيّرتين X وY في حالة من الحالات الثلاث التي يقدّمها الشكل 41 .

العلاقة الوظيفية أو العاملية.في حالة الظاهرة التي يمثّلها الشكل 41a ، يوجد علاقة عاملية متبادلة بين قيم المتغيّرتين Y وX : لكلّ قيمة x_i نخصّص قيمة محدّدة y_i ،

وبالعكس . المتوسّط الشرطي $\bar{y}_i$ المتعلّق بـ x_i يساوي y_i ؛ كذلك ، المتوسّط الشرطي $\bar{x}_i$ المتعلّق بـ y_i يساوي x_i : يتطابق منحنيا الانحدار عندها مع منحنى العلاقة القائمة .

إذاً ، يحكم قانون دقيق العلاقات بين المتغيّرتين . وغالباً ما نصادف هذا الأمر في مجال الفيزياء ، مثلاً عند حرارة ثابتة ، يرتبط ضغط كتلة غاز معيّنة P وحجمها V بواسطة العلاقة العاملية التالية :

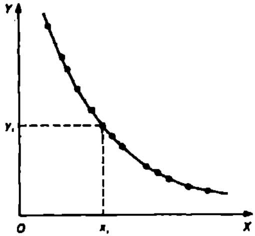
$$P.V = k$$

حيث k هي ثابتة .

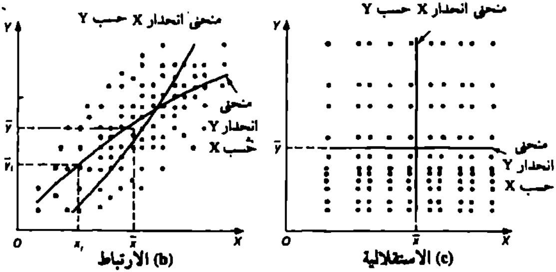
الإستقلالية : بالمقابل ، يمثّل الشكل 41c حالة الاستقلالية بين المتغيّرتين X وY : تتطابق توزيعات كلّ من المتغيّرتين الشرطية مع التوزيع الهامشي المناسب وتكون بالتالي متطابقة فيما بينها (أنظر الفصل I ، ص 50) . نستنتج أنّه لكلّ من المتغيّرتين ، تتساوى المتوسّطات الشرطية فيما بينها وتساوي أيضاً المتوسّط الهامشي :

$$\bar{x}_j = \bar{x} \qquad \bar{y}_i = \bar{y} .$$

إذاً ، يكون منحنيا الانحدار خطّين متوازيين مع محوري الإحداثيات : في حالة الاستقلالية ، لا تعطينا معرفة قيمة إحدى المتغيّرتين، X مثلاً، أي معلومات إضافية حول توزيع المتغيّرة الأخرى ، وبصورة خاصة عن قيمتها المتوسّطة .



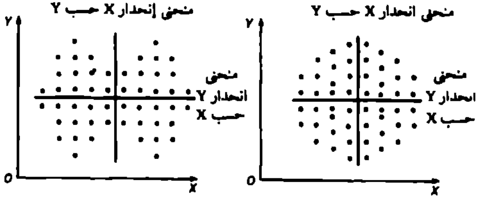
(a) علاقة عاملية متبادلة



الشكل 41

هكذا فإن مستندات المصارف الفرنسية المالية وانتاج الأرزّ في اليابان هما كمّيتان مستقلّتان : لا تعطينا معرفة انتاج الأرزّ في اليابان أي معلومات حول قيمة المستندات المالية ، والعكس بالعكس .

ملاحظة : إذا كانت الاستقلالية تعني وجود خطّي انحدار متوازيين مع محوري الإحداثيات ، فالعكس ليس صحيحاً : الحصول على خطّيّ انحدار متوازيين مع المحورين لا يعني بالضرورة أنّ المتغيّرتين موضع الدراسة هما مستقلّتان . تحدّد الاستقلالية ، في الواقع ، بالتطابق الحاصل بين التوزيعات الشرطية . إلاّ أنّه قد يوجد توزيعات لها نفس المتوسّط دون أن تكون متطابقة : بصورة خاصّة ، قد تكون تشتّاتها مختلفة . في هذه الحالة ، نتكلّم عن غياب متبادل للارتباط ، وليس عن الاستقلالية (الشكل 42) . لكن على الصعيد العملي لا يختلف الظرفان كثيراً بشكل عام : ففي كلتي الحالتين ، لا تعطينا معرفة إحدى المتغيرتين أيّة معلومات إضافية حول قيمة المتغيّرة الأخرى المتوسّطة .



الشكل 42 . مثلان حول غياب الارتباط المتبادَل

الارتباط : الوضع الذي يصفه الشكل 41b هو وضع وسيط بين الحالتين القصويين السابقتين . بما أنّ منحنى انحدار Y حسب X هو غير متواز مع المحور السيني فإنّ معرفة القيمة التي تأخذها X تأتي بفكرة إضافية حول القيم التي تأخذها Y : إذا كانت $X=x_1$ ، فإنّ Y تأخذ بالمتوسّط القيمة $\bar{y}_1$ ، وليس $\bar{y}$. نقول أنّ Y هي في ارتباط مع X .

كذلك ، بما أنّ منحنى انحدار X حسب Y ليس متوازياً مع المحور OY ، فإنّ X هي في ارتباط مع Y .

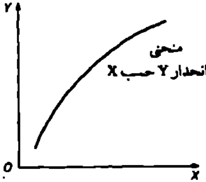
إذاً ، دون أن يتحكّم قانون دقيق بعلاقاتهما ، يوجد نوع من التبعية بين المتغيّرتين المدروستين . تتكرّر هذه الحالة بكثرة ، لا سيّما في مجال الاقتصاد وإدارة الأعمال وعلى العموم في مجال العلوم الإنسانية .

هكذا يظهر لنا فحص الشكل 40 أنّ الراتب الشهري للموظّفين هو في ارتباط مع العمر . توجد علاقة معيّنة بين هاتين الكمّيتين ، بمعنى أنّه ، حتّى السنّ 55 عاماً ، يتزايد الراتب المتوسط مع العمر . لكن هذه العلاقة ليست إلزامية : فقد يربح بعض الموظّفين الشباب أكثر من بعض الموظّفين الأكبر سنّاً .

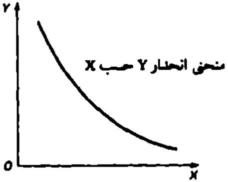
عندما تتّجه تغيّرات ظاهرتين في نفس الإتّجاه ، نقول أن الارتباط هو مباشر أو إيجابي (الشكل 43) .

عندما يكون اتّجاها التغيّرات متعاكسين ، نقول أنّ الارتباط هو عكسي أو سلبي (الشكل 44) .

عندما يكون منحنيا الانحدار خطّين غير متوازيين مع محوري الإحداثيات ، يوجد ارتباط خطّي .



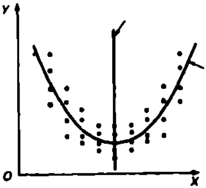
الشكل 43 . ارتباط إيجابي



الشكل 44 . ارتباط سلبي

ملاحظة : بخلاف الاستقلالية ، الارتباط ليس خاصّة متبادَلة : قد تكون Y مرتبطة مع X دون أن تكون X مرتبطة مع Y (الشكل 45) .

باختصار ، عندما تكون المتغيّرة Y في ارتباط مع المتغيّرة X ، يسمح لنا منحنى انحدار Y حسب X بتلخيص العلاقة الموجودة بين المتغيّرتين بشكل ملائم . وتزداد أهمّية هذا التلخيص كلّما كان تمثيل منحنى الإنحدار لمجمل توزيع المتغيّرتين « صادقاً » أكثر ، بعبارة أخرى كلّما كانت النقاط (x_i, y_i) مركزة أكثر حول منحنى الانحدار . وتقاس كثافة العلاقة هذه بواسطة نسبة الارتباط .



الشكل 45 . الارتباط ليس خاصّة متبادلة

2 . نسبة الارتباط

كما سبق أن أثبتنا (القسم I ، ص 170) يُساوي تباين المتغيّرة Y الهامشي حاصل جمع عنصرين : تباين المتوسّطات الشرطية $\bar{y}_i$ ومتوسّط التباينات الشرطية $V_i(Y)$:

$$V(Y) = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2 + \frac{1}{n}\sum_{i=1}^{k} n_{i.}\, V_i(Y) .$$

العنصر الأوّل :

$$V(\bar{y}_i) = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2$$

يعبّر عن قسم التباين الهامشي المفسّر بتغيّر المتوسّطات الشرطية $\bar{y}_i$ ، أي بمنحنى انحدار Y حسب X .

بالمقابل ، فإنّ العنصر الثاني :

$$\frac{1}{n}\sum_{i=1}^{k} n_{i.}\, V_i(Y)$$

يقيس قسم التباين الهامشي الذي ينتج عن تشتّت النقاط (x, y) حول منحنى الانحدار : إنّه التباين المتبقّي ، الذي لا يفسّره الانحدار .

بوسعنا إذاً أن نكتب :

يساوي التباين الكلّي (التباين الهامشي) حاصل جمع التباين المفسّر بالانحدار مع التباين المتبقّي .

ويستند تعريف نسبة الارتباط إلى هذه التجزئة .

A . تعريف

يساوي مربّع نسبة الارتباط خارج قسمة التباين المفسّر بالانحدار على التباين الكلّي :

$$\eta^2 = \frac{\text{التباين المفسّر}}{\text{التباين الكلي}} = 1 - \frac{\text{التباين المتبقّي}}{\text{التباين الكلي}}$$

بالتالي ، يساوي مربّع نسبة ارتباط Y مع X :

$$\eta^2_{Y/X} = \frac{V(\bar{y}_i)}{V(Y)} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{k} n_{i.} V_i(Y)}{V(Y)}$$

ونعرّف بنفس الطريقة نسبة ارتباط X مع Y :

$$\eta^2_{X/Y} = \frac{V(\bar{x}_j)}{V(X)} = 1 - \frac{\frac{1}{n}\sum_{j=1}^{l} n_{.j} V_j(X)}{V(X)}$$

بشكل عام ، تكون قيمتا نسبتي الارتباط مختلفتين .

B . الخصائص

1 . في غياب ارتباط Y مع X ، يكون منحنى انحدار Y حسب X خطاً متوازياً مع المحور السيني : تتساوى كلّ المتوسّطات الشرطية $\bar{y}_i$ فيا بينها وتساوي أيضاً المتوسّط الهامشي $\bar{y}$. إذاً ، يكون تباين المتوسّطات الشرطية :

$$V(\bar{y}_i) = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2$$

مساوياً لصفر ونسبة الارتباط $\eta_{Y/X}$ مساوية لصفر أيضاً .

2 . في حالة العلاقة العاملية بين Y وX ، يكون كلّ من التباينات الشرطية :

$$V_i(Y) = \frac{1}{n_{i.}} \sum_{j=1}^{l} n_{ij}(y_j - \bar{y}_i)^2 , \qquad i = 1, 2, \ldots, k$$

مساوياً لصفر لأنّ كلّ النقاط التي تمثّل التوزيع توجد على منحنى الانحدار :

$$y_j = \bar{y}_i .$$

إذاً ، متوسّط هذه التباينات الشرطية ، أي التباين المتبقّي ، يساوي صفراً هو أيضاً بينما تساوي نسبة الارتباط $\eta_{Y/X}$ واحداً .

3 . عند وجود ارتباط بين Y وX ، تقترب نسبة الارتباط $\eta_{Y/X}$ من 1 كلّما كانت حصّة التباين المفسّر بالانحدار من التباين الكلّي أكبر ، بعبارة أخرى كلّما كانت درجة الارتباط أقوى .

إذاً ، تشكّل نسبة الارتباط $\eta_{Y/X}$ قياساً لكثافة علاقة متغيّرة معيّنة Y مع متغيّرة أخرى X . وهي تحقّق عدم المساواة التالية :

$$0 \leqslant \eta_{Y/X} \leqslant 1 .$$

عندما تكون مساوية لصفر ، فهذا يعني غياب ارتباط Y مع X .
عندما تكون مساوية لواحد ، فهذا يعني وجود علاقة عاملية .

بين هذين الحالتين القصويين ، تكون كثافة علاقة Y مع X أقوى كلّما اقتربت قيمة نسبة الارتباط أكثر من 1 . ويحكم خصائص التباين ، هذه القيمة ، كما سنرى لاحقاً ، هي ثابتة بالنسبة لاستبدال نقطة الأصل والوحدة : إنّها عدد لا بعد له .

بما أنّ نسبة الارتباط لا تستدعي قياس متغيّرة العلاقة ، يمكن استعمالها لوصف كثافة علاقة متغيّرة كمّية مع متغيّرة نوعية ، كما بالنسبة لعلاقة متغيّرتين نوعيّتين .

بالمقابل ، من سيئاتها أنّها تتعلّق بعدد فئات أو كيفيّات متغيّرة العلاقة قيمتها تكبر بشكل عام مع قيمة هذا العدد .

C . حساب نسبة الارتباط عملياً

ويتمّ ذلك انطلاقاً من قاعدة التعريف :

$$\eta^2_{Y/X} = \frac{V(\bar{y}_i)}{V(Y)} .$$

أي بوضع قيمتي تباين المتوسّطات الشرطية $V(\bar{y}_i)$ والتباين الهامشي V(Y) مكانهما :

$$\eta^2_{Y/X} = \frac{\frac{1}{n}\sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2}{\frac{1}{n}\sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2} = \frac{\sum_{i=1}^{k} n_{i.}(\bar{y}_i - \bar{y})^2}{\sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2}.$$

وكذلك

$$\eta^2_{X/Y} = \frac{V(\bar{x}_j)}{V(X)} = \frac{\sum_{j=1}^{l} n_{.j}(\bar{x}_j - \bar{x})^2}{\sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2}.$$

عملياً ، كي نحسب نسبة الارتباط ، نستعمل الطرق الموضوعة لتسهيل حساب التباين : القاعدة المبسّطة، واستبدال المتغيّرة .

مثلاً . لقد قمنا برسم (الشكل 40 ، ص 173) منحنى انحدار R حسب A الذي يخصّ توزيع الموظّفين حسب الراتب الشهري R والعمر A .

لنحسب نسبة ارتباط R حسب A .

بما أنّ الدخل هو متغيّرة متواصلة ، جُمّعت المشاهدات في فئات . عند الحسابات ، نأخذ كمتغيّرة إحصائية مركز كلّ فئة r_i [1] .

لتسهيل الحسابات ، قمنا في هذا المثل باستبدال المتغيّرة التالي :

$$r'_i = \frac{r_i - r_0}{a} = \frac{r_i - 1\,100}{50}.$$

انطلاقاً من قاعدة التعريف :

$$\eta^2_{R/A} = \frac{V(\bar{r}_j)}{V(R)}.$$

لقد جمّعنا الحسابات في الجدول 18 .

---

(1) باستثناء الفئتين الطرفين ، المفتوحتين ، حيث نأخذ متوسّط الحصة المفترض .

الجدول 18 . توزيع الموظّفين حسب الراتب الشهري والعمر .
جدول حساب منحنى انحدار ونسبة ارتباط R حسب A .

			1	2	3	4	5	6	7	8				
		فئات العمر	أقل من 25	من 25 إلى 30	من 30 إلى 35	من 35 إلى 40	من 40 إلى 45	من 45 إلى 50	من 50 إلى 55	55 سنة وأكثر				
	فئات الرواتب	a_j / r_i	20,0	27,5	32,5	37,5	42,5	47,5	52,5	60,0	حواصل $n_{i.}$ (1)	r'_i (2)	$n_{i.} r'_i$ (3)=(1).(2)	$n_{i.} r'^2_i$ (4)=(3).(2)
1	أقل من 800 F	700	207	121	38	17	10	2	7	3	405	− 8	− 3 240	+ 25 920
2	من 800 إلى 900 F	850	302	461	513	103	86	6	10	2	1 483	− 5	− 7 415	+ 37 075
3	من 900 إلى 1 000 F	950	18	526	682	367	613	431	105	60	3 002	− 3	− 9 006	+ 27 018
4	من 1 000 إلى 1 200 F	1 100		111	342	298	416	400	226	37	1 910	0	0	
5	من 1 200 إلى 1 500 F	1 350		1	3	102	227	263	98	10	792	+ 5	+ 3 960	+ 19 800
6	من 1 500 إلى 2 000 F	1 750				10	22	13	12	5	70	+ 13	+ 910	+ 11 830
7	أكثر من 2000F	2 200				1	14	6	7	5	33	+ 22	+ 726	+ 15 972
		حواصل $n_{.j}$ (1)	527	1 220	1 578	1 106	1 380	1 201	465	130	7 695		− 14 065	+ 137 615
		$\sum_i n_{ij} r'_i$ (2)	− 3 220	− 4 046	− 4 900	− 1 106	− 620	+ 277	+ 379	+ 51	− 14 065		$\sum_i n_{i.} r'_i$	$\sum_i n_{i.} r'^2_i$
		$\bar{r}'_j$ (3)=$\frac{(2)}{(1)}$	− 6,11	− 3,97	− 3,11	− 1,00	− 0,45	+ 0,23	+ 0,82	+ 0,39				
		$\bar{r}'_j \sum_i n_{ij} r'_i$ (4)=(3).(2)	19 674,20	19 238,62	15 239,00	1 106,00	279,00	63,71	310,78	19,89	56 011,28	$\sum_j n_{.j} \bar{r}'^2_j$		

ـ متوسّط وتباين R

$$1. \qquad \bar{r}' = \frac{1}{n} \sum_i n_i r'_i = \frac{-14\,065}{7\,695} = -1{,}83$$

إذاً :

$$\bar{r} = a\bar{r}' + r_0 = 50 \times (-1{,}83) + 1\,100 = 1\,008{,}5 .$$

$$2. \qquad V(R') = \frac{\sum_i n_{i.} r'^2_i - n\bar{r}'^2}{n}$$

$$= \frac{137\,615{,}00 - (-14\,065) \times (-1{,}83)}{7\,695} = \frac{111\,876{,}05}{7\,695} .$$

إذاً

$$V(R) = a^2\ V(R') = (50)^2 \times \frac{111\,876{,}05}{7\,695} = 36\,347$$

$$\sigma_R = \sqrt{36\,347} = 190{,}7 .$$

ـ المتوسّطات الشرطية $\bar{r}_j$ وتباين المتوسّطات الشرطية $V(\bar{r}_j)$

بحكم استبدال المتغيّرة الذي أجريناه : $r_i = ar'_i + r_0$ يوجد بين كلّ من المتوسّطين الشرطيين $\bar{r}_j$ و $\bar{r}'_j$ نفس العلاقة القائمة بين المتغيّرتين r_i و r'_i (أنظر « الإحصاء الوصفي » ، الفصل الخامس ، القسم I ، الفقرة 3.B) :

$$\bar{r}_j = a\bar{r}'_j + r_0 , \qquad j = 1, 2, \ldots, l .$$

بالتالي يوجد بين تبايني المتوسطين الشرطيين $V(\bar{r}_j)$ و $V(\bar{r}'_j)$ العلاقة التالية (أنظر «الاحصاء الوصفي» ، الفصل الخامس ، القسم II) :

$$V(\bar{r}_j) = a^2\ V(\bar{r}'_j) .$$

1 . المتوسّطات الشرطية $\bar{r}'_j$

كُرّست الأسطر من (1) إلى (3) من الجدول لحساب المتوسّطات الشرطية $\bar{r}'_j$.

نحصل على السطر (2) بجمعنا ، في كلّ عامود من الجدول ، حواصل الضرب $n_{ij}\, r'_i$. مثلاً :

$$\sum_i n_{i6}\, r'_i = 2 \times (-8) + 6 \times (-5) + 431 \times (-3) + 480 \times 0$$
$$+ 263 \times 5 + 13 \times 13 + 6 \times 22 = +277 .$$

حاصل جمع هذا السطر :

$$\sum_j \sum_i n_{ij}\, r'_i = \sum_i \left(\sum_j n_{ij}\right) r'_i = \sum_i n_{i.}\, r'_i$$

يساوي حاصل جمع العامود (3) ، ما يعطي ، على هذا الصعيد ، وسيلة لمراقبة دقّة الحسابات .

نحصل على السطر (3) بقسمتنا عنصراً عنصراً السطر (2) على السطر (4) ، والسطر (3) يعطي متوسّطات R′ الشرطية المتعلّقة بـ A .

$$\bar{r}_j = \frac{\sum_i n_{ij} r'_i}{n_{.j}}$$

ويسمح بالتالي برسم منحنى انحدار R حسب A . هكذا :

$$\bar{r}_1 = 50 \times (-6{,}11) + 1\,100 = 794{,}5$$
$$\bar{r}_2 = 50 \times (-3{,}97) + 1\,100 = 901{,}5$$

الخ . . .

لقد تمّ بهذه الطريقة حساب عامود « الراتب الشهري المتوسّط » من الجدول 16 ، ص 166 .

2 . تباين المتوسّطات الشرطية $V(\bar{r}_j)$

انطلاقاً من قاعدة التباين المتبسّطة :

$$V(\bar{r}_j) = \frac{\sum_j n_{.j}\bar{r}_j^2 - n\bar{r}^2}{n}.$$

كُرّس السطر (4) من الجدول لحساب $\sum_j n_{.j}\bar{r}_j^2$. نحصل عليه بضربنا ، عنصراً عنصراً ، السطر (2) بالسطر (3) . انطلاقاً من تعريف المتوسّط الشرطي ، عناصر السطر (2) تساوي :

$$\sum_i n_{ij} r'_i = n_{.j}\bar{r}_j.$$

إذاً ، عناصر السطر (4) هي :

$$\bar{r}_j \sum_i n_{ij} r'_i = n_{.j}\bar{r}_j^2$$

ويساوي حاصل جمع هذا السطر :

$$\sum_j n_{.j}\bar{r}_j^2.$$

لدينا إذاً :

$$V(\bar{r}_j) = \frac{\sum_j n_{.j}\bar{r}_j^2 - n\bar{r}^2}{n}$$

$$= \frac{56\,011{,}20 - (-14\,065 \times (-1{,}83))}{7\,695} = \frac{30\,272{,}25}{7\,695}$$

إذاً :

$$V(\overline{r}_j) = a^2\, V(\overline{r}_j') = (50)^2 \times \frac{30\,272{,}25}{7\,695} = 9\,835$$

$$\sigma_{\overline{r}_j} = \sqrt{9\,835} = 99{,}2\,.$$

ـ نسبة ارتباط R مع A

$$\eta^2_{R/A} = \frac{V(\overline{r}_j)}{V(R)} = \frac{a^2\, V(\overline{r}_j')}{a^2\, V(R')} = \eta^2_{R'/A}\,.$$

بالتالي

$$\eta^2_{R/A} = \frac{30\,272{,}25}{111\,876{,}05} = 0{,}27\,.$$

نستنتج إذاً أنّه لحساب نسبة ارتباط R حسب A ، يكفي أن نحسب نسبة ارتباط R′ حسب A : نسبة الارتباط هي ثابتة بالنسبة لاستبدال نقطة الأصل والوحدة .

ـ تجزئة التباين الهامشي

لقد رأينا (القسم I ، ص 170) أنّه يمكننا تجزئة التباين الهامشي V(R) فيصبح مجموع عنصرين : تباين المتوسّطات الشرطية $V(\overline{r}_j)$ ومتوسّط التباينات الشرطية $V_j(R)$. يمثّل العنصر الأوّل حصّة التباين المفسّر بالانحدار ويمثّل العنصر الثاني التباين المتبقّي :

$$V(R) = V(\overline{r}_j) + \frac{1}{n}\sum_j n_j\, V_j(R)$$

= V مفسّر + V متبقّي

لقد سمحت لنا الحسابات التي أجريناها في الجدول 18 بتحديد قيمتي V(R) و$V(\overline{r}_j)$. يمكننا الحصول على التباين المتبقّي ، الذي يقتضي حسابه حساب كلّ من التباينات الشرطية $V_j(R)$ ، بالطرح . لدينا :

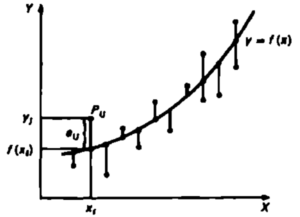
V(R) = V مفسّر + V متبقّي

36 347 = 9 835 + 26 512 .

هكذا ، في مثلنا هذا ، يفسّر منحنى الانحدار 27% فقط من تباين الرواتب ، هذا ما يبيّنه مربّع نسبة الارتباط . هذه القيمة صغيرة : ارتباط الراتب مع العمر هو نسبياً ضعيف .

3 . مبدأ طريقة المربّعات الصغرى

يملك منحنى انحدار Y حسب X خاصّة جديرة بالملاحظة : فبالنسبة لهذا المنحنى يكون مجموع مربّعات الانحرافات (الفروقات) ، المقاسة بالتوازي مع المحور الصادي ، بين النقاط الملحوظة P_{ij} والمنحنى ، حدّاً أدنى (أصغر) (الشكل 46) .



الشكل 46 . منحنى المربّعات الصغرى

لنأخذ المنحنى ذا المعادلة : $y = f(x)$.

إنّ مجموع مربّعات الانحرافات e_{ij} ، مقاسة بالتوازي مع المحور الصادي ، بين كلّ من النقاط الملحوظة P_{ij} والمنحنى ، المجموع المرجّح ، عند الاقتضاء ، بالمقادير n_{ij} التي تناسب كلاً من النقاط ، يساوي :

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} e_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} [y_j - f(x_i)]^2 .$$

يمكننا تجزئة هذا المجموع بالطريقة التالية :

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{i.} \frac{n_{ij}}{n_{i.}} [y_j - f(x_i)]^2 = \sum_{i=1}^{k} n_{i.} \sum_{j=1}^{l} \frac{n_{ij}}{n_{i.}} [y_j - f(x_i)]^2 ,$$

حيث :

$$\frac{n_{ij}}{n_{i.}} = f_{j/i}$$

تمثّل تردّد y_j الشرطي متعلّقة بـ x_i .

لقيمة x_i مثبّتة ، $f(x_i)$ هي عدد ثابت . إذاً المجموع :

$$S_i = \sum_{j=1}^{l} \frac{n_{ij}}{n_{i.}} [y_j - f(x_i)]^2 = \frac{1}{n_{i.}} \sum_{j=1}^{l} n_{ij} [y_j - f(x_i)]^2$$

يساوي متوسّط مربّعات الانحرافات (الفروقات) بين قيم المتغيّرة الشرطية $Y/(x=x_i)$ الملحوظة وهذا العدد الثابت .

عند دراستنا لخصائص المتوسّط الحسابي الجبرية ، أظهرنا (الكتاب الأوّل ، الفصل VI ، القسم I ، الفقرة 3.B) أنّ متوسّط مربّعات الانحرافات بين القيم الملحوظة y_j للمتغيّرة الإحصائية وعدد ثابت y_0 ، يساوي مجموع عنصرين :

$$\frac{1}{n}\sum_j n_j(y_j - y_0)^2 = \frac{1}{n}\sum_j n_j(y_j - \bar{y})^2 + (\bar{y} - y_0)^2$$
$$= V(Y) + (\bar{y} - y_0)^2 .$$

إذا طبّقنا هذه النتيجة على توزيع Y الشرطي متعلّقة بـ x_i ، وبما أنّ $f(x_i)$ عدد ثابت ، نحصل على :

$$S_i = \frac{1}{n_{i.}}\sum_{j=1}^{l} n_{ij}[y_j - f(x_i)]^2 = V_i(Y) + [\bar{y}_i - f(x_i)]^2 .$$

بالتالي :

$$S = \sum_{i=1}^{k} n_{i.} S_i = \sum_{i=1}^{k} n_{i.} V_i(Y) + \sum_{i=1}^{k} n_{i.}[\bar{y}_i - f(x_i)]^2 .$$

يكون هذا المجموع حدّاً أدنى (أصغر : minimum) إذا كان عنصره الثاني يساوي صفراً ، أي عندما يكون ، لكلّ x_i :

$$f(x_i) = \bar{y}_i .$$

هكذا فالمنحنى $g=f(x)$ ، حيث يكون مجموع مربّعات الانحرافات ، مقاسة بالتوازي مع المحور الصادي ، حدّ أدنى ، هو منحنى انحدار Y حسب X [1] . منحنى الانحدار هو إذن منحنى المربّعات الصغرى ، أي نوعاً ما المنحنى الأقرب من النقاط التي تمثّل التوزيع .

تسمح هذه الخاصّة بتحديد منحنى Y حسب X عندما نعرف مسبقاً شكله التحليلي ، وذلك بطريقة أسهل من الطريقة المعروضة سابقاً . لنفترض مثلاً أنّ هذا المنحنى هو خطّ مستقيم معادلته :

---

(1) هذه الخاصّة هي نتيجة مباشرة من الخاصّة التي أثبتناها في الكتاب الأوّل ، الفصل X ، القسم I ، الفقرة 3.B ، إذ بالنسبة للمتوسّط الحسابي يكون مجموع مربّعات الانحرافات حدّاً أدنى . ويحقّق منحنى انحدار Y حسب X هذه الخاصّة لكلّ القيم x_i .

$$y = ax + b .$$

سيتمّ تقدير قيمتي المتغيّرين الوسيطين a وb بشكل يكون فيه مجموع مربّعات الانحرافات ، مقاسة كما أشرنا سابقاً ، حدّاً أدنى .

إنّ البحث عن قيمة المتغيّرات الوسيطية لمنحنى انحدار نفترض أنّنا نعرف شكله التحليلي مسبقاً ، يطلق عليه اسم تسوية المنحنى مع التوزيع الملحوظ. والطريقة التي تقوم على تحقيق هذه التسوية بشكل يكون فيه مجموع مربّعات انحرافات النقاط الملحوظة عن المنحنى حدّاً أدنى هي طريقة المربّعات الصغرى .

القسم III

التسوية الخطّية

1 . التسوية الخطّية على طريقة المربّعات الصغرى : A . حالة المشاهدات المفردة ؛ B . حالة المشاهدات المجمّعة في فئات ؛ C . تحويلات بسيطة تسمح ببسط استعمال التسوية الخطية .ـ 2 . مُعامِل الارتباط الخطّي . A . تعريف ؛ B . الخصائص ؛ C . الحساب العملي .ـ 3 . خصائص خطوط التسوية : A . المواضع الخاصّة بخطوط المربّعات الصغرى ؛ B . استعمال خط التسوية في التقدير والتوقّع ؛ C . تجزئة التباين الهامشي .

تلعب التسوية الخطّية دوراً مميّزاً في التحليل وتوقّع الظواهر الإقتصادية : تحليل الاستهلاك ، توقّع الطلب ، الخ . إنّ معظم النماذج الاقتصادية المترية التي تسعى ، مثلاً ، إلى تمثيل تطوّر استهلاك بعض المواد تبعاً لتطوّر المداخيل والأسعار ، هي نماذج خطّية .

قد يبدو استعمال الرسومات الخطّية لتمثيل نماذج اقتصادية معقّدة تبسيطاً تعسّفياً للحقيقة . إلاّ أنّه في حالات عديدة ، ما عدا بعض تحويلات الكمّيات المدروسة ـ لا سيّما التحويل اللوغاريتمي ـ يظهر اعتماد دالّة خطّية ، عملياً ، كفرضية معقولة . إذ غالباً ما تكون المعطيات التي بحوزتنا غير دقيقة لتجعل من التمثيلات الأكثر تعقيداً والتي لا يكون تبريرها النظري دوماً متيناً أمراً وهمياً . لهذا السبب تجعلنا بساطة الحسابات التي تؤدّي إليها التسوية الخطّية نفضّلها عن أي شكل آخر للتسوية .

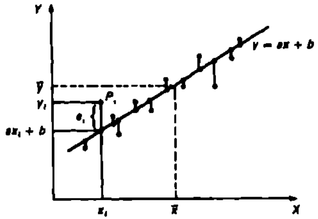
1 . التسوية الخطّية على طريقة المربّعات الصغرى

لنأخذ توزيع متغيّرتين X وY نفترضهما مسبقاً في ارتباط خطّي : منحنيا انحدار

Y حسب X وX حسب Y هما خطّان مستقيمان . تقوم تسوية خطّ انحدار Y حسب X على طريقة المربّعات الصغرى على تبنّي ، من بين كلّ خطوط المسطّح ، الخطّ الذي يجعل مجموع مربّعات الانحرافات بين النقاط الملحوظة وبينه ، مقاسة بالتوازي مع المحور الصادي ، حداً أدنى . إنّه الخطّ حيث المسافة إلى النقاط التمثيلية ، محدّدة كمجموع مربّعات الانحرافات ، هي أصغر ما يمكن .

A . حالة المشاهدات المفرّدة

عندما تكون المشاهدات مفرّدة ، كلّ وحدة إحصائية يناسبها زوج القيم (x_i, y_i) ممثّلاً بالنقطة p_i (الشكل 47) .



الشكل 47 . خطّ المربّعات الصغرى

أ ـ معادلة خطّ المربّعات الصغرى

لنأخذ الخطّ ذا المعادلة : $y = ax + b$

ولنحسب قيمة انحرافات النقاط الملحوظة عن الخط ، مقاسة بالتوازي مع المحور الصادي :

$$e_i = y_i - ax_i - b\,, \qquad i = 1, 2, \ldots, n\,.$$

مجموع مربّعات هذه الانحرافات يساوي :

$$S = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2\,.$$

إنّ خط المربّعات الصغرى يطابق قيمتي المعاملين a وb اللتين تجعلان هذه

الكمّية حدّاً أدنى ، نحصل على هذا الحدّ الأدنى إذا جعلنا مشتقّي S الجزئيّتين بالنسبة لِـ a وb تساويان صفراً .

لنبحث أوّلاً ، بالنسبة لِـ a مثبّتة ، عن قيمة b التي تجعل S حدّاً أدنى :

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{n}(y_i - ax_i - b) = 0$$

($\frac{\partial S}{\partial b}$ هي تفاضل S بالنسبة لِـ b) .

بالتالي :

$$\sum_{i=1}^{n} y_i - a\sum_{i=1}^{n} x_i - nb = 0$$

وإذا قسمنا على n عنصري هذه المعادلة :

$$\frac{1}{n}\sum_{i=1}^{n} y_i - a.\frac{1}{n}\sum_{i=1}^{n} x_i - b = 0$$

$$\bar{y} - a\bar{x} - b = 0$$

أو :

$$\bar{y} = a\bar{x} + b .$$

توضّح هذه العلاقة أنّ خطّ المربّعات الصغرى يمرّ بالنقطة الوسط $(\bar{x}, \bar{y})$.

لنضع قيمة b التي وجدناها : $b = \bar{y} - a\bar{x}$ ، مكانها في عبارة S :

$$S = \sum_{i=1}^{n}[y_i - ax_i - (\bar{y} - a\bar{x})]^2 = \sum_{i=1}^{n}[(y_i - \bar{y}) - a(x_i - \bar{x})]^2 .$$

هكذا نحصل على قيمة حدّ S الأدنى ، حيث a مثبّتة . لنبحث الآن عن قيمة a التي تجعل هذه الكمّية حدّاً أدنى :

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{n}(x_i - \bar{x})[(y_i - \bar{y}) - a(x_i - \bar{x})] = 0 .$$

عند التوسيع :

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) - a\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0 .$$

إذاً ، قيمة ميل (pente) خطّ المربّعات الصغرى هي :

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

أي ، بناء على تعريفي التباين والتغاير (cov) (أنظر القسم I) :

$$a = \frac{\text{cov}(XY)}{\sigma_X^2}.$$

بالمختصر :

خطّ تسوية Y حسب X : y = ax + b يمرّ بالنقطة الوسط $(\bar{x}, \bar{y})$ وميله هو :

$$a = \frac{\text{cov}(XY)}{\sigma_X^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}.$$

كذلك ، يمرّ خطّ تسوية X حسب $x = a'y + b'$ بالنقطة الوسط $(\bar{x}, \bar{y})$ وميله هو :

$$a' = \frac{\text{cov}(XY)}{\sigma_Y^2} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (y_i - \bar{y})^2}.$$

ب ـ حساب خطّ المربّعات الصغرى عملياً

لحساب مُعامَلَيْ خطّ التسوية ، نعتمد الطرق المستعملة لتبسيط حساب التباين والتغاير : القواعد المتبسّطة واستبدالات نقطة الأصل (أنظر القسم I ، ص 167) .

مثلاً . يعرض الجدول 19 تطوّرات الإنتاج المحلّي الإجمالي P والاستهلاك C خلال السنوات من 1960 إلى 1969 . يظهر لنا الرسم البياني (الشكل 48) أنّ النقاط التمثيلية تظهر على نفس الخطّ تقريباً .

لنسوِّ خطّي الانحدار على طريقة المربّعات الصغرى . كي نسهّل الحسابات ، عمدنا في هذا المثل إلى استبدالي نقطتي الأصل :

$$P_i' = P_i - P_0 = P_i - 460, \qquad C_i' = C_i - C_0 = C_i - 280.$$

تمّ تجميع الحسابات في الجدول 20 .

ـ المتوسّطات ، التباينات والتغاير

$$1. \qquad \bar{P'} = \frac{1}{n}\sum_{i=1}^{n} P_i' = \frac{-86}{10}, \qquad \bar{C'} = \frac{1}{n}\sum_{i=1}^{n} C_i' = \frac{-47}{10}$$

إذاً :

$$\bar{P} = \bar{P}' + P_0 = -8{,}6 + 460 = 451{,}4\,,$$
$$\bar{C} = \bar{C}' + C_0 = -4{,}7 + 280 = 275{,}3\,.$$

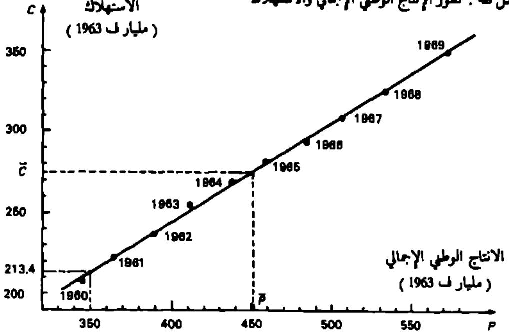
الجدول 19 . تطوّر الانتاج الوطني الإجمالي والاستهلاك من 1960 إلى 1969 .

المصدر : المحاسبة الوطنية

الوحدة : مليار فرنك 1963

السنة	الانتاج الوطني الإجمالي	الاستهلاك
1960	346	209
1961	365	222
1962	390	238
1963	412	255
1964	439	269
1965	460	281
1966	486	294
1967	508	309
1968	533	326
1969	575	350

الشكل 48 . تطوّر الإنتاج الوطني الإجمالي والاستهلاك



الجدول 20 . تطوّر الإنتاج الوطني الإجمالي والاستهلاك الفردي

جدول الحسابات

i (1)	P_i (2)	C_i (3)	P'_i (4)	C'_i (5)	P'^2_i (6) = (4).(4)	C'^2_i (7) = (5).(5)	$P'_i C'_i$ (8) = (4).(5)
1	346	209	− 114	− 71	12 996	5 041	8 094
2	365	222	− 95	− 58	9 025	3 364	5 510
3	390	238	− 70	− 42	4 900	1 764	2 940
4	412	255	− 48	− 25	2 304	625	1 200
5	439	269	− 21	− 11	441	121	231
6	460	281	0	+ 1	0	1	0
7	486	294	26	+ 14	676	196	364
8	508	309	48	+ 29	2 304	841	1 392
9	533	326	73	+ 46	5 329	2 116	3 358
10	575	350	115	+ 70	13 225	4 900	8 050
المجموع	—	—	− 86	− 47	51 200	18 969	31 139
			$\sum_{i=1}^{10} P'_i$	$\sum_{i=1}^{10} C'_i$	$\sum_{i=1}^{10} P'^2_i$	$\sum_{i=1}^{10} C'^2_i$	$\sum_{i=1}^{10} P'_i C'_i$

2. $$V(P') = \frac{\sum_i P'^2_i - n\bar{P}'^2}{n} = \frac{51\,200 - (-86) \times (-8{,}6)}{10} = \frac{50\,460{,}4}{10}$$

$$V(C') = \frac{\sum_i C'^2_i - n\bar{C}'^2}{n} = \frac{19\,969 - (-47) \times (-4{,}7)}{10} = \frac{18\,748{,}1}{10}$$

إذاً :

$$V(P) = V(P') = 5\,046{,}04\,, \qquad V(C) = V(C') = 1\,874{,}81$$

$$\sigma_P = \sqrt{5\,046{,}04} = 70{,}0\,, \qquad \sigma_C = \sqrt{1\,874{,}81} = 43{,}3\,.$$

3. $$\operatorname{cov}(P'\,C') = \frac{\sum_i P'_i\,C'_i - n\bar{P}'\,\bar{C}'}{n} = \frac{31\,139 - (-86) \times (-4{,}7)}{10} = \frac{30\,734{,}8}{10}$$

إذاً :

$$\operatorname{cov}(PC) = \operatorname{cov}(P'\,C') = 3\,073{,}48\,.$$

- خطّ تسوية C حسب P

يمرّ هذا الخطّ ذو المعادلة : C = aP + b بالنقطة الوسط $(\bar{P},\ \bar{C})$.

ميله يساوي :

$$a_{C/P} = \frac{\operatorname{cov}(PC)}{\sigma_P^2} = \frac{\operatorname{cov}(P'C')}{\sigma_{P'}^2} = a_{C'/P'}$$

$$= \frac{30\,734{,}8}{50\,460{,}4} = 0{,}609 .$$

معادلة خطّ تسوية C حسب P هي :

$$C - \overline{C} = 0{,}61(P - \overline{P})$$

$$C = 0{,}61\ P + \overline{C} - 0{,}61\ \overline{P} = 0{,}61\ P - 0{,}1 .$$

عملياً ، يمرّ هذا الخط بنقطة الأصل . بما أنّ هذه النقطة لا تظهر على الرسم البياني ، كي نرسم الخطّ نحسب نقطة أخرى ، مثلاً :

$$P = 350 , \qquad C = 213{,}4 .$$

- خطّ تسوية P حسب C

هذا الخطّ ذو المعادلة $P = a'C + b'$ يمرّ أيضاً بالنقطة الوسط $(\overline{C}, \overline{P})$.

ميله يساوي :

$$a'_{P/C} = \frac{\operatorname{cov}(PC)}{\sigma_C^2} = \frac{\operatorname{cov}(P'C')}{\sigma_{C'}^2} = a'_{P'/C'} = \frac{30\,734{,}8}{18\,748{,}1} = 1{,}639 .$$

معادلة خطّ تسوية P حسب C هي :

$$P - \overline{P} = 1{,}64(C - \overline{C})$$

$$P = 1{,}64\ C + \overline{P} - 1{,}64\ \overline{C} = 1{,}64\ C + 0{,}1 .$$

كي نخطّه على الرسم البياني ، نكتب معادلته بالشكل :

$$C - \overline{C} = \frac{1}{1{,}64}(P - \overline{P})$$

$$C = 0{,}61\ P - 0{,}1$$

إذاً خطّا التسوية هما عملياً متطابقان .

B . حالة المشاهدات المجمّعة في فئات

عندما تكون المشاهدات مجمّعة في فئات ، نأخذ بشكل عام كمتغيّرات إحصائية ، عند الحسابات ، مراكز كلّ فئة x_i و y_i . هكذا نفترض أنّ المشاهدات مجمّعة

في المركز P_{ij} للمستطيلات المحدّدة بأزواج فسحات الفئات (الشكل 49) . إذاً كلّ نقطة P_{ij} ، إحداثياها (x_i, y_j) ، يناسبها المقدار n_{ij} .

أ ـ معادلة خطّ المربّعات الصغرى

يجري تحديد خطّ التسوية تماماً بنفس طريقة حالة المشاهدات المفرَدة ، ولكن الدلالات معقّدة أكثر بفعل المقادير n_{ij} المنسوبة لكلّ نقطة .

لنأخذ الخط ذا المعادلة : $y = ax + b$

ولنحسب قيم انحرافات النقاط الملحوظة P_{ij} عن الخطّ ، مقاسة بالتوازي مع المحور الصادي :

$$e_{ij} = y_j - ax_i - b , \qquad i = 1, 2, \ldots, k ; j = 1, 2, \ldots, l .$$

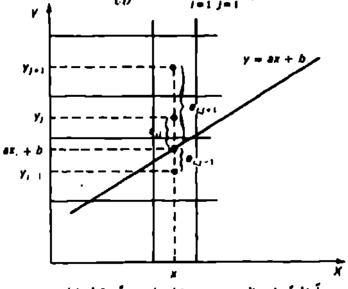
إنّ مجموع مربّعات الانحرافات ، مرجحاً بالمقادير n_{ij} المخصّصة لكلّ من النقاط ، يساوي :

$$S = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij} e_{ij}^2 = \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}(y_j - ax_i - b)^2$$

إنّ خطّ تسوية Y حسب X ، على طريقة المربّعات الصغرى ، يطابق قيمتي المعاملين a وb اللتين تجعلان هذه الكمّية حدّاً أدنى . ونحصل على هذا الحدّ الأدنى عندما نساوي بالصفر مشتقّي S الجزئيتين بالنسبة لـ a وb .

لنبحث أوّلاً ، لقيمة معطيّة لـ a ، عن قيمة b التي تجعل S حدّاً أدنى :

$$\frac{\partial S}{\partial b} = -2 \sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}(y_j - ax_i - b) = 0$$



الشكل 49 . خطّ المربّعات الصغرى . مشاهدات مجمّعة في فئات

بالتالي :

$$\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\, y_j - a\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\, x_i - b\sum_{i=1}^{l}\sum_{j=1}^{k} n_{ij} = \sum_{j=1}^{l} n_{.j}\, y_j - a\sum_{i=1}^{k} n_{i.}\, x_i - nb = 0 \quad (1)$$

وذلك لأنّ :

$$\sum_{i=1}^{k} n_{ij} = n_{.j} \qquad \sum_{j=1}^{l} n_{ij} = n_{i.} \qquad \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = n\,.$$

إذا قسمنا على n عنصري المعادلة (1) ، نحصل على :

$$\frac{1}{n}\sum_{j=1}^{l} n_{.j}\, y_j - a \times \frac{1}{n}\sum_{i=1}^{k} n_{i.}\, x_i - b = 0$$

$$\overline{y} - a\overline{x} - b = 0\,.$$

إذاً ، يمرّ خطّ المربّعات الصغرى بالنقطة الوسط $(\overline{x}, \overline{y})$.

لنضع قيمة b التي وجدناها مكانها في عبارة S :

$$S = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}[y_j - ax_i - (\overline{y} - a\overline{x})]^2$$

$$= \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}[(y_j - \overline{y}) - a(x_i - \overline{x})]^2\,.$$

لنبحث الآن عن قيمة a التي تجعل هذه الكمّية حداً أدنى :

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \overline{x})\,[(y_j - \overline{y}) - a(x_i - \overline{x})]^2 = 0\,.$$

عند التوسيع :

$$\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \overline{x})(y_j - \overline{y}) - a\sum_{i=1}^{k} n_{i.}(x_i - \overline{x})^2 = 0\,.$$

إذاً ميل خطّ تسوية Y حسب X على طريقة المربّعات الصغرى هو :

$$a = \frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \overline{x})(y_j - \overline{y})}{\sum_{i=1}^{k} n_{i.}(x_i - \overline{x})^2}$$

أي ، بناء على تعريفي التباين والتغاير (أنظر القسم I) :

$$a = \frac{\text{cov}\,(XY)}{\sigma_X^2}$$

بالمختصر : إنّ خط تسوية Y حسب X ، $y = ax + b$ يمرّ بالنقطة الوسط $(\bar{x}, \bar{y})$ ، وميله هو :

$$a = \frac{\mathrm{cov}\,(XY)}{\sigma_X^2} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2}.$$

كذلك ، يمرّ خط تسوية X حسب Y، $x = a'y + b'$ بالنقطة الوسط $(\bar{x}, \bar{y})$ وميله هو :

$$a' = \frac{\mathrm{cov}\,(XY)}{\sigma_Y^2} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y})}{\sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2}.$$

ب ـ حساب خطّ المربّعات الصغرى عملياً

عملياً ، كي نحسب معاملي خط التسوية ، نستعمل طرق القواعد المبسّطة واستبدالات المتغيّرة المعتمدة لتسهيل حساب التباين والتغاير (أنظر القسم I ، ص 167) .

مثلاً . جرى استقصاء على 2000 أسرة وأعطى النتائج المشار إليها في الجدول 21 في ما يتعلّق بتوزيع الدخل والاستهلاك الكلّي .

يسمح لنا التمثيل البياني (الشكل 50) بافتراض وجود علاقة خطّية .

لنحدّد عن طريقة المربّعات الصغرى خطّي التسوية .

نجري الحسابات آخذين كمتغيّرات إحصائية مراكز الفئات . وقد تمّ تحديد مركزي الفئتين الطرفين اصطلاحياً[1] .

لتسهيل الحسابات عمدنا في هذا المثل إلى استبدالَيْ المتغيّرة :

$$C_i' = \frac{C_i - C_0}{\alpha} = \frac{C_i - 1\,100}{50}, \qquad R_j' = \frac{R_j - R_0}{\beta} = \frac{R_j - 1\,100}{100}$$

وقد تمّ تجميع الحسابات في الجدول 22 .

(1) نختار قيمة قريبة من متوسّط الحصة المفترض .

الجدول 21 . توزيع عيّنة من 2000 أسرة حسب دخلها واستهلاكها الكلّي

المداخيل / الاستهلاك	أقلّ من 800F	من 800 إلى أقلّ من 1000F	من 1000 إلى أقلّ من 1200F	من 1200 إلى أقلّ من 1600F	من 1600 إلى أقلّ من 2000F	2000F وأكثر	المجموع
أقلّ من 800F	178	141	58				337
من 800 إلى أقلّ من 1000F	43	567	98	17			725
من 1000 إلى أقلّ من 1200F		57	320	38			415
من 1200 إلى أقلّ من 1500F			19	165	16	23	223
من 1500 إلى أقلّ من 1800F				80	76	18	174
1800F وأكثر					36	50	86
المجموع	221	765	495	300	128	91	2000

ـ المتوسّطات ، التباينات والتغاير

1. $$\overline{C}' = \frac{1}{n}\sum_{i=1}^{k} n_{i.}\, C_i' = \frac{-\ 1\ 339}{2\ 000} = -\ 0{,}669\ 5\ ,$$

$$\overline{R}' = \frac{1}{n}\sum_{j=1}^{l} n_{.j}\, R_j' = \frac{+\ 656}{2\ 000} = +\ 0{,}328\ 0$$

إذاً :

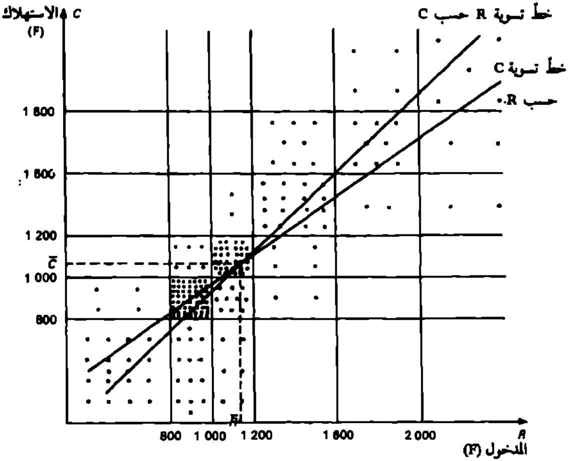
$$\overline{C} = \alpha\overline{C}' + C_0 = 50 \times (-\ 0{,}669\ 5) + 1\ 100 = 1\ 066{,}5$$

$$\overline{R} = \alpha\overline{R}' + R_0 = 100 \times 0{,}328\ 0 + 1\ 100 = 1\ 132{,}8\ .$$

$$V(C') = \frac{\sum_i n_{i.}\, C_i'^2 - n\overline{C}'^2}{n}$$

$$= \frac{90\ 221 - (-\ 1\ 339) \times (-\ 0{,}699\ 5)}{2\ 000} = \frac{89\ 324{,}535\ 5}{2\ 000}$$

$$V(R') = \frac{\sum_j n_{.j}\, R_j' - n\overline{R}'^2}{n} = \frac{33\ 404 - 215{,}168\ 0}{2\ 000} = \frac{33\ 188{,}832\ 0}{2\ 000}$$



الشكل 50 . توزيع عيّنة من 2000 أسرة حسب مدخولها واستهلاكها الكلّي

إذاً :

$$V(C) = \alpha^2\, V(C') = \frac{(50)^2 \times 89\,324{,}535\,5}{2\,000} = 111\,655{,}67$$

$$V(R) = \beta^2\, V(R') = \frac{(100)^2 \times 33\,188{,}832\,0}{2\,000} = 165\,944{,}16$$

$$\sigma_C = \sqrt{111\,655{,}67} = 334{,}1\,, \qquad \sigma_R = \sqrt{165\,944{,}16} = 407{,}4\,.$$

3. $$\operatorname{cov}(R'\,C') = \frac{\sum_i \sum_j n_{ij}\, C'_i\, R'_j - n\overline{C'}\,\overline{R'}}{n}.$$

خُصّص العامودان (5) و(6) من جدول الحسابات لحساب العبارة :

$$\sum_i \sum_j n_{ij}\, C'_i\, R'_j\,.$$

الجدول 22 . توزيع عيّنة من 2000 أسرة حسب مدخولها واستهلاكها الكلّي . جدول حساب خطّي الانحدار ومعامل الارتباط الخطّي . (لقراءة من اليسار إلى اليمين) .

	فئات المداخيل	أقل من 800F	من 800 إلى 1 000 F	من 1 000 إلى 1 200 F	من 1 200 إلى 1 600 F	من 1 600 إلى 2 000 F	2 000 F وأكثر						
Classes de consommation	C_i \ R_j	700	900	1 100	1 400	1 800	2 500	حواصل الجمع n_i (1)	C (2)	$n_i C_i$ (3)=(1).(2)	$n_i C_i^2$ (4)=(3).(2)	$\sum_j n_{ij} R_j$ (5)	$C_i \sum_j n_{ij} R_j$ (6)=(5).(2)
أقل من 800 F	700	178	141	58				377	− 8	− 3 016	24 128	− 994	+ 7 952
من 800 إلى 1 000 F	900	43	567	98	17			725	− 4	− 2 900	11 600	− 1 255	+ 5 020
من 1 000 إلى 1 200 F	1 100		57	320	38			415	0	0	0	0	0
من 1 200 إلى 1 500 F	1 350			19	165	16	23	223	+ 5	+ 1 115	5 575	+ 929	+ 4 645
من 1 500 إلى 1 800 F	1 650				80	76	18	174	+ 11	+ 1 914	21 054	+ 1 024	+ 11 264
1 800 F وأكثر	2 000					36	50	86	+ 18	+ 1 548	27 864	+ 952	+ 17 136
	حواصل n_j (1)	221	765	495	300	128	91	2 000		− 1 339	90 221	+ 656	+ 46 017
										$\sum_i n_i C_i$	$\sum_i n_i C_i^2$		$\sum_i \sum_j n_{ij} C_i R_j$

R_j (2)	− 4	− 2	0	+ 3	+ 7	+ 14		
$n_j R_j$ (3)=(1).(2)	− 884	− 1 530	0	+ 900	+ 896	+ 1 274	+ 656	$\sum_j n_j R_j$
$n_j R_j^2$ (4)=(3).(2)	3 536	3 060	0	2 700	6 272	17 836	33 404	$\sum_j n_j R_j^2$
$\sum_i n_{ij} C_i$ (5)	− 1 596	− 3 396	− 764	+ 1 637	+ 1 564	+ 1 213	− 1 339	
$R_j \sum_i n_{ij} C_i$ (6)=(5).(2)	+ 6 384	+ 6 792	0	+ 4 911	+ 10 948	+ 16 982	+ 46 017	$\sum_i \sum_j n_{ij} C_i R_j$

نحصل على العامود (5) بجمعنا ، في كلّ سطر من الجدول ، حواصل الضرب $n_{ij} R_j'$ ، مثلاً :

$$\sum_j n_{1j} R_j' = 178 \times (-4) + 141 \times (-2) + 58 \times 0 = -994$$

مجموع هذا العامود :

$$\sum_i \sum_j n_{ij} R'_j = \sum_j \left(\sum_i n_{ij}\right) R'_j = \sum_j n_{.j} R'_j$$

يساوي مجموع السطر (3) ، ما يعطي ، على هذا الصعيد ، وسيلة ممكنة لمراقبة دقّة الحسابات .

نحصل على العامود (6) بضربنا ، عنصراً عنصراً ، العامود (5) بالعامود (2) . مجموعه يساوي العبارة التي نبحث عنها .

يمكننا إجراء نفس الحساب ، بطريقة مماثلة ، إنطلاقاً من السطرين (5) و(6) من الجدول .

نحصل على :

$$\text{cov}(R'C') = \frac{46\,017 - (-1\,339) \times 0{,}328\,0}{2\,000} = \frac{46\,456{,}192\,0}{2\,000}$$

إذاً :

$$\text{cov}(RC) = \alpha\beta\,\text{cov}(R'C') = (50 \times 100).\frac{46\,456{,}192\,0}{2\,000} = 116\,140{,}48\,.$$

- خطّ تسوية C حسب R

هذا الخطّ ذو المعادلة $C = aR + b$ يمرّ بالنقطة الوسط $(\bar{R}, \bar{C})$، ميله يساوي :

$$a_{C/R} = \frac{\text{cov}(RC)}{\sigma^2_R} = \frac{\alpha}{\beta}\frac{\text{cov}(R'C')}{\sigma^2_{R'}} = \frac{\alpha}{\beta} a_{C'/R'}$$

$$= \frac{50}{100}.\frac{46\,456{,}192\,0}{33\,188{,}832\,0} = 0{,}699\,8\,.$$

إذاً معادلة خطّ تسوية C حسب R هي :

$$C - \bar{C} = 0{,}70(R - \bar{R})$$

$$C = 0{,}70\,R + \bar{C} - 0{,}70\,\bar{R} = 0{,}70\,R + 273{,}6\,.$$

كي نرسم هذا الخطّ ، نحسب نقطة أخرى ، مثلاً :

$$R = 2\,000\,, \qquad C = 1\,673{,}6\,.$$

. خطّ تسوية R حسب C

هـذا الخطّ ذو المعادلة $R = a'C + b'$ يمرّ أيضاً بالنقطة الوسط $(\overline{C}, \overline{R})$ وميله يساوي :

$$a'_{R/C} = \frac{\operatorname{cov}(RC)}{\sigma_C^2} = \frac{\beta}{\alpha}\frac{\operatorname{cov}(R'C')}{\sigma_{C'}^2} = \frac{\beta}{\alpha}a'_{R'/C'}$$

$$= \frac{100}{50}\cdot\frac{46\,456{,}192\,0}{89\,324{,}535\,5} = 1{,}040\,2\,.$$

إذاً ، معادلة خطّ تسوية R حسب C هي :

$$R - \overline{R} = 1{,}04(C - \overline{C})$$

$$R = 1{,}04\,C + \overline{R} - 1{,}04\,\overline{C} = 1{,}04\,C + 23{,}6\,.$$

كي نخطّه على الرسم البياني ، نكتب معادلته بالشكل :

$$C - \overline{C} = \frac{1}{1{,}04}(R - \overline{R})\,, \qquad C = 0{,}96\,R - 22{,}7\,.$$

ونحسب نقطة غير النقطة الوسط ، مثلاً :

$$R = 2\,000\,, \qquad C = 1\,897{,}3\,.$$

C . تحويلات بسيطة تسمح بسط استعمال التسوية الخطّية

في عدد معيّن من الحالات ، يمكننا ردّ دراسة العلاقة بين متغيّرتين إلى دراسة تسوية خطّية ، وذلك بواسطة تحويلات بسيطة . لقد صادفنا بعض الأمثلة عند دراستنا للمقاييس الوظيفية (« الإحصاء الوصفي » ، الفصل IV) .

1 . المخطّط الأسّي

لنأخذ كمّيتين x وy تربطهما العلاقة التالية :

$$y = y_0\,a^x\,. \qquad (1)$$

إنّ هذه العلاقة (وهي الوظيفة أو الدّالة الأسّية) تمثّل الظواهر حيث يكون معدّل تغيّر y بالنسبة لـ x ثابتاً :

$$\frac{dy/y}{dx} = k\,. \quad (k \text{ ثابتة })\,.$$

غالباً ما يكون هذا المخطّط ملائماً لوصف تطوّر (تصاعدياً أو تنازلياً) كمّية معيّنة تبعاً للوقت .

لنأخذ لوغاريتم عنصري العبارة (1) :

$$\log y = \log y_0 + x \log a$$

إذا وضعنا :

$$Y = \log y , \qquad \alpha = \log a , \qquad \beta = \log y_0 ,$$

نحصل على :

$$Y = \alpha x + \beta .$$

إذاً ، تمثّل العلاقة (1) بخطّ مستقيم على رسم بياني نصف لوغاريتمي (واحد من المحورين هو بقياسٍ لوغاريتمي) ، ويمكننا تسوية هذا الخطّ على النقاط الملحوظة (x_i, Y_i) على طريقة المربّعات الصغرى .

2 . مخطّط ذو مرونة ثابتة

لنأخذ كمّيتين x وy تربط بينهما العلاقة التالية

$$y = y_0 \, x^{\alpha} . \qquad (2)$$

إنّ هذه العلاقة (دالّة أو وظيفة القوّة) تمثّل الظواهر حيث تكون مرونة y بالنسبة لِـ x ثابتة :

$$\frac{dy/y}{dx/x} = \alpha . \qquad (\alpha \text{ ثابتة })$$

غالباً ما يُستعمل هذا المخطّط ، مثلاً ، لوصف تطوّر الاستهلاك تبعاً للدخل أو للأسعار (وظيفة الاستهلاك) أو تطوّر الإنتاج تبعاً للعمل أو لرأس المال (وظيفة الإنتاج) .

لنأخذ لوغاريتم عنصري العبارة (2) :

$$\log y = \log y_0 + \alpha \log x .$$

إذا وضعنا :

$$Y = \log y , \qquad X = \log x , \qquad \beta = \log y_0$$

نحصل على :

$$Y = \alpha X + \beta .$$

إذاً ، تمثّل العلاقة (2) بخطّ مستقيم على رسم بياني بمحورين لوغاريتميّين ، ويمكننا تسوية هذا الخطّ على النقاط الملحوظة (X_i, Y_i) على طريقة المربّعات الصغرى .

3 . المخطّط الغوسيّ

لقد رأينا (الفصل III ، ص 121) أنّه يوجد بين قيمة متغيّرة إحصائية موزّعة طبيعياً x وتردّدها (تكرارها) المتراكم y ، العلاقة التالية :

$$y = \Pi\left(\frac{x - m}{\sigma}\right) \qquad (3)$$

لنأخذ التحويل المعاكس :

$$\Pi^{-1}(y) = \frac{x - m}{\sigma} .$$

إذا وضعنا :

$$t = \Pi^{-1}(y)$$

(حيث t هي ، تعريفاً ، المتغيّرة الطبيعية المركزة المختصرة) ، نحصل على :

$$t = \frac{1}{\sigma} x - \frac{m}{\sigma} .$$

إذاً ، تُمثّل العلاقة (3) بخطّ مستقيم ، نسمّيه خطّ هنري ، على رسم بياني غوسيّ - حسابي . ويمكننا تقدير متغيري القانون الطبيعي (المعتدل) m و σ بواسطة تسوية هذا الخطّ على النقاط الملحوظة (x_i, t_i) .

إنّ استعمال تحويلات من هذا النوع يزيد حتماً من حقل تطبيق التسوية الخطّية .

2 . معامل الارتباط الخطّي

يهدف معامل الارتباط الخطّي إلى قياس كثافة العلاقة الخطّية بين المتغيّرتين X وY .

A . تعريف

نعرّف مُعامِل الارتباط الخطّي r بين X وY كخارج القسمة التالي :

$$r = \frac{\text{cov}(XY)}{\sigma_X \sigma_Y} .$$

بناءً على تناظر هذا التعريف ، يميّز معامل الارتباط الخطّي كثافة علاقة Y حسب X وعلاقة X حسب Y على السواء .

يوجد بين معامل الارتباط الخطي وميلي خطّيْ التسوية العلاقتان التاليتان :

ـ خطّ تسوية Y حسب X :

$$a = \frac{\text{cov}\,(XY)}{\sigma_X^2} = r\frac{\sigma_Y}{\sigma_X}, \qquad r = a\frac{\sigma_X}{\sigma_Y}.$$

ـ خطّ تسوية X حسب Y :

$$a' = \frac{\text{cov}\,(XY)}{\sigma_Y^2} = r\frac{\sigma_X}{\sigma_Y}, \qquad r = a'\frac{\sigma_Y}{\sigma_X}$$

B . الخصائص

1 ـ إذا كانت المتغيّرتان X وY مستقلّتين ، فإنّ معامل الارتباط الخطّي يساوي صفراً .

في الحقيقة ، عندما تكون المتغيّرتان مستقلّتين (أنظر الفصل I ، ص 62) : $\text{cov}\,(XY) = 0$ ، إذن :

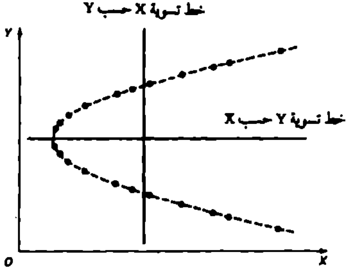
$$r = \frac{\text{cov}\,(XY)}{\sigma_X\,\sigma_Y} = 0\,.$$

إلاّ أنّ : $r = 0$ لا تعني بالضرورة الاستقلالية بين X وY ، فقط تشير إلى أنّ خطّي التسوية هما متوازيان مع محوري الإحداثيات . في الواقع ، إذا كانت $\sigma_X \neq 0$ و $\sigma_Y \neq 0$ ،

$a = r\sigma_Y/\sigma_X = 0$ عندما يكون $r = 0$

$a' = r\sigma_X/\sigma_Y = 0$ عندما يكون $r = 0$.

هكذا، على الشكل 51 لا يوجد استقلالية بين X وY ، بل علاقة عاملية . لكن خطّي التسوية يوازيان المحورين و$r = 0$. هذا المثل يُظهر أنّ معامل الارتباط الخطّي لا يجب أن يُستعمل لوصف كثافة الارتباط إلاّ حيث يكون هذا الارتباط تقريباً خطّياً .



الشكل 51

2 . إنّ معامل الارتباط الخطّي محصور بين 1 - و1+ :

$$-1 \leqslant r \leqslant +1$$

لنأخذ المتغيّرتين المركزتين :

$$x' = x - \bar{x}, \qquad y' = y - \bar{y}$$

والعبارة :

$$\frac{1}{n}\sum_i\sum_j n_{ij}(\lambda x'_i + y'_j)^2 \qquad (1)$$

حيث λ هي عدد معيّن . لدينا :

$$\frac{1}{n}\sum_i\sum_j n_{ij}(\lambda x'_i + y'_j)^2 = \lambda^2 \frac{1}{n}\sum_i n_{i.} x'^2_i + 2\lambda \frac{1}{n}\sum_i\sum_j n_{ij} x'_i y'_j + \frac{1}{n}\sum_j n_{.j} y'^2_j$$
$$= \lambda^2 \sigma_X^2 + 2\lambda \operatorname{cov}(XY) + \sigma_Y^2$$

إلّا أنّه ، مهما تكن λ ، العبارة (1) هي إيجابية أو تساوي صفراً . وتكون قيمة مثلّث الحدود ذي الدرجة الثانية (حسب λ) ، :

$$a\lambda^2 + b\lambda + c.$$

حيث $a = \sigma_X^2$ هي كمّية إيجابية ، إيجابية أو تساوي صفراً ، مهما تكن λ ، إذا كان مميّزه سلبياً أو يساوي صفراً بالتالي :

$$\Delta' = [\operatorname{cov}(XY)]^2 - \sigma_X^2 \sigma_Y^2 \leqslant 0 \quad (^1)$$

(1) يُعرف عدم المساواة هذا باسم عدم مساواة شوارتز (Schwartz)

إذاً :

$$r^2 - 1 = \frac{[\text{cov}(XY)]^2}{\sigma_X^2 \sigma_Y^2} - 1 \leq 0$$

و :

$$-1 \leq r \leq +1$$

3 . إذا ربطت بين المتغيّرتين X وY علاقة عاملية خطّية ، فإنّ معامل الارتباط الخطّي يساوي 1 – أو 1 + .

لنأخذ العلاقة العاملية : $y_i = ax_i + b$

لدينا :

$r = +1$ إذا كان $a > 0$ (علاقة مباشرة) .

$r = -1$ إذا كان $a < 0$ (علاقة غير مباشرة) .

لنكتب في الواقع أنّ خطّ العلاقة يمرّ بالنقطة الوسط $(\bar{x}, \bar{y})$:

$$y_i - \bar{y} = a(x_i - \bar{x}).$$

بالتالي :

$$\text{cov}(XY) = \frac{1}{n}\sum_i n_i(x_i - \bar{x})(y_i - \bar{y}) = a\frac{1}{n}\sum_i n_i(x_i - \bar{x})^2 = a\sigma_X^2$$

$$\sigma_Y^2 = \frac{1}{n}\sum_i n_i(y_i - \bar{y})^2 = a^2\frac{1}{n}\sum_i n_i(x_i - \bar{x})^2 = a^2\sigma_X^2$$

وبما أنّ σ_Y إيجابي فهو يساوي :

$+ a\sigma_X$ إذا كان $a > 0$

$- a\sigma_X$ إذا كان $a < 0$

أي :

$$\sigma_Y = |a|\sigma_X.$$

بالتالي :

$$r = \frac{\text{cov}(XY)}{\sigma_X \sigma_Y} = \frac{a\sigma_X^2}{\sigma_X |a| \sigma_X} = \frac{a}{|a|}$$

إذاً :

$r = 1$ إذا كان $a > 0$

$r = -1$ إذا كان $a < 0$

4 . بين هاتين الحالتين القصويين ، غياب الارتباط والعلاقة العاملية

الخطّية ، يمثّل معامل الارتباط مقياساً للتبعية الخطّية ، على درجاتها المتفاوتة ، بين متغيّرتين إحصائيتين . وتقترب قيمته المطلقة من 1 كلّما كانت هذه التبعية أقوى : سوف نرى ، في الواقع ، في الفقرة التالية أن مربّع مُعامِل الارتباط يمثّل قسم التباين الكلّي المفسّر بخطّ التسوية .

يكون معامل الارتباط الخطّي إيجابياً في حالة العلاقة المباشرة ، وسلبياً في حالة العلاقة العكسية . ولا معنى له ، إذاً لا ينبغي استعماله ، إلّا في الحالة حيث يمكننا اعتبار العلاقة بين المتغيّرتين تقريباً خطّية .

كما سنرى لاحقاً ، معاملُ الارتباط الخطّي هو كمّية ثابتة بالنسبة لتغيير نقطة الأصل والوحدة : إنّه عدد لا بعد له .

C . حساب مُعامِل الارتباط الخطّي عملياً

مثل 1.المشاهدات المفرّدة

لنعد إلى دراسة العلاقة بين الإنتاج الوطني الإجمالي P والاستهلاك الفردي من 1960 إلى 1969 (أنظر ص 190) .

لتسهيل الحسابات ، المعروضة في الجدول 20 ، ص 192 ، عمدنا إلى تغيير نقطتي الأصل التالي :

$$P_i' = P_i - 460\,, \qquad C_i' = C_i - 280\,.$$

انطلاقاً من تعريف مُعامِل الارتباط :

$$r = \frac{\operatorname{cov}(PC)}{\sigma_P\,\sigma_C}$$

$$\operatorname{cov}(PC) = \operatorname{cov}(P'C') = \frac{\sum_{i=1}^{n} P_i' C_i' - n\overline{P'}\,\overline{C'}}{n} = \frac{30\,734{,}8}{10}$$

$$\sigma_P^2 = \sigma_{P'}^2 = \frac{\sum_{i=1}^{n} P_i'^2 - n\overline{P}'^2}{n} = \frac{50\,460{,}4}{10}$$

$$\sigma_C^2 = \sigma_{C'}^2 = \frac{\sum_{i=1}^{n} C_i'^2 - n\overline{C}'^2}{n} = \frac{18\,748{,}1}{10}$$

إذاً :

$$r = \frac{30\,734{,}8}{\sqrt{50\,460{,}4 \times 18\,748{,}1}} = \frac{30\,734{,}8}{30\,757{,}7}$$

في هذا المثل ، يقترب معامل الارتباط الخطّي كثيراً من 1 ، ما يعني تقريباً وجود علاقة عاملية خطّية مباشرة بين المتغيّرتين . وبالفعل ، لقد أظهر حساب خطّي التسوية أنّهما تقريباً متطابقان.

ملاحظة . في حالة مثل هذه ، حيث المشاهدات مفرّدة (مشاهدة واحدة في السنة) ، لم يكن بالإمكان حساب نسبة الارتباط التي تستدعي تجميع المشاهدات في فئات : فعدد هذه المشاهدات ليس كبيراً بشكل كاف . بالمقابل ، يمكن دائماً حساب معامل الارتباط الخطّي .

مثل 2 . المشاهدات المجمّعة في فئات

لنعد الآن إلى تحليل توزيع الدخل والاستهلاك الكلّي انطلاقاً من نتائج الإستقصاء الذي أجري على 2000 أسرة (أنظر ص 196) .

لتسهيل الحسابات المعروضة في الجدول 22 ، ص 199 ، عمدنا إلى استبدال المتغيّرات التالي :

$$C'_i = \frac{C_i - C_0}{\alpha} = \frac{C_i - 1\,100}{50}, \qquad R'_j = \frac{R_j - R_0}{\beta} = \frac{R_j - 1\,100}{100}.$$

إنطلاقاً من تعريف معامل الارتباط :

$$r = \frac{\operatorname{cov}(RC)}{\sigma_R\,\sigma_C}$$

$$\operatorname{cov}(RC) = \alpha\beta \operatorname{cov}(R'\,C') = \alpha\beta\,\frac{\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\,C'_i\,R'_j - n\overline{C}'\,\overline{R}'}{n} = \alpha\beta\,\frac{46\,456{,}192\,0}{2\,000}$$

$$\sigma_C^2 = \alpha^2\,\sigma_{C'}^2 = \alpha^2\,\frac{\sum_{i=1}^{k} n_{i.}\,C_i'^2 - n\overline{C}'^2}{n} = \alpha^2\,\frac{89\,324{,}535\,5}{2\,000}$$

$$\sigma_R^2 = \beta^2\,\sigma_{R'}^2 = \beta^2\,\frac{\sum_{j=1}^{l} n_{.j}\,R_j'^2 - n\overline{R}'^2}{n} = \beta^2\,\frac{33\,188{,}812\,0}{2\,000}$$

$$r = \frac{\operatorname{cov}(RC)}{\sigma_C\,\sigma_R} = \frac{\alpha\beta \operatorname{cov}(R'\,C')}{\alpha\sigma_{C'}\cdot\beta\sigma_{R'}} = \frac{\operatorname{cov}(R'\,C')}{\sigma_{C'}\,\sigma_{R'}}$$

$$r = \frac{46\,456{,}9}{\sqrt{89\,324{,}54 \times 33\,188{,}83}} = 0{,}85.$$

نقرّ إذن أنّه للحصول على معامل ارتباط X وY ، يكفي حساب معامل ارتباط X′ وY′ ، لا يتغيّر معامل الارتباط الخطّي عند تغيير نقطة الأصل والوحدة .

3 ـ خصائص خطوط التسوية

A . المواضع الخاصّة بخطوط المربّعات الصغرى

إنّ خطّي تسوية Y حسب X وX حسب Y يمرّان بالنقطة الوسط $(\bar{x}, \bar{y})$ للتوزيع . معادلتاهما :

ـ بالنسبة لخطّ تسوية Y حسب X :

$$y - \bar{y} = a(x - \bar{x}), \qquad (1)$$

ـ بالنسبة لخطّ تسوية X حسب Y :

$$x - \bar{x} = a'(y - \bar{y})$$

أيْ ، في نفس نظام المحاور :

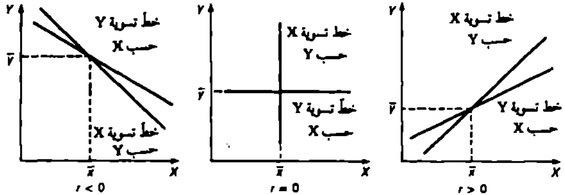
$$y - \bar{y} = \frac{1}{a'}(x - \bar{x}). \qquad (2)$$

إذا وضعنا مكان a وa′ عبارتيهما تبعاً لِـ r ، σ_x و σ_y ، نحصل على التوالي على :

$$(1) \rightarrow y - \bar{y} = r\frac{\sigma_y}{\sigma_x}(x - \bar{x})$$

$$(2) \rightarrow y - \bar{y} = \frac{1}{r}\frac{\sigma_y}{\sigma_x}(x - \bar{x}).$$

إذن ، لميلي الخطّين نفس الإشارة الجبرية ، إشارة r . بالقيمة المطلقة ، ميل خطّ تسوية X حسب Y هو دائماً أكبر من ميل خطّ تسوية Y حسب X لأنّ قيمة r المطلقة هي أصغر من 1 (الشكل 52) .



الشكل 52 . المواضع الخاصّة بخطوط المربّعات الصغرى

في حالة الاستقلالية ، يكون الخطّان موازيين لمحوري الإحداثيات ومتعامدين فيما بينهما (يشكّلان زاوية قائمة فيما بينهما) . وتتناقص زاوية الخطّين تدريجياً كلّما ازدادت قيمة r المطلقة . عندما تصبح $|r|$ مساوية لـ 1 ، يتطابق الخطّان ويوجد علاقة عاملية خطّية بين المتغيّرتين X و Y .

B . استعمال خطّ التسوية في التقدير والتوقّع

عند غياب أيّة معلومات أخرى ، أفضل تقدير يمكن إجراؤه للقيمة المجهولة التي تأخذها متغيّرة إحصائية معيّنة Y هو متوسّطها $\bar{y}$.

بالمقابل ، إذا كانت Y على ارتباط مع متغيّرة أخرى X ، فإنّ معرفة قيمة هذه الأخيرة تسمح بتحسين تقدير Y . ضمن الفرضية أنّ هذه العلاقة هي خطّية ، معادلة خطّ المربّعات الصغرى هي :

$$y - \bar{y} = r\frac{\sigma_Y}{\sigma_X}(x - \bar{x}) .$$

لقيمة x_0 تأخذها X نقدّر Y بـ :

$$y^* = \bar{y} + r\frac{\sigma_Y}{\sigma_X}(x_0 - \bar{x}) .$$

تسمح الفكرة الإضافية التي يعطيها وجود العلاقة الخطّية ومعرفة قيمة X بزيادة التصحيح $+ r\frac{\sigma_Y}{\sigma_X}(x_0 - \bar{x})$ إلى التقدير الأصلي $\bar{y}$.

في الحالة الأولى ، يشكّل قياس تشتّت القيم الملحوظة y_j حول القيمة المقدّرة ،

$$V(Y) = \frac{1}{n}\sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2 ,$$

مؤشّر انحراف بين التوقّعات والتحقيقات .

في الحالة الثانية ، يتألّف هذا المؤشّر من متوسّط مربّعات انحرافات القيم الملحوظة عن خطّ التسوية :

$$V_R = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}\left[(y_j - \bar{y}) - r\frac{\sigma_Y}{\sigma_X}(x_i - \bar{x})\right]^2$$

هذه الكمّية هي حدّ أدنى بناء على تعريف خطّ المربّعات الصغرى . لنحسب قيمة هذا الحدّ الأدنى :

$$V_R = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(y_j - \bar{y})^2 - 2r\frac{\sigma_Y}{\sigma_X}\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(y_j - \bar{y})(x_i - \bar{x})$$
$$+ r^2\frac{\sigma_Y^2}{\sigma_X^2}\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})^2 .$$

إلاّ أنّ :

$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2 = \sigma_X^2$$
$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(y_j - \bar{y})^2 = \frac{1}{n}\sum_{j=1}^{l} n_{.j}(y_j - \bar{y})^2 = \sigma_Y^2$$
$$\frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}(x_i - \bar{x})(y_j - \bar{y}) = \text{cov}(XY) = r\sigma_X\sigma_Y$$

إذاً :

$$V_R = \sigma_Y^2 - 2r\frac{\sigma_Y}{\sigma_X}r\sigma_X\sigma_Y + r^2\frac{\sigma_Y^2}{\sigma_X^2}\sigma_X^2 .$$

أخيراً نحصل على :

$$V_R = (1 - r^2)\,V(Y) .$$

تم إذن ، بالمتوسّط ، اختصار (تصغير) الانحراف بين التوقّعات والتحقيقات ، في خارج القسمة التالي :

$$\frac{V(Y) - V_R}{V(Y)} = \frac{V(Y) - (1 - r^2)\,V(Y)}{V(Y)} = r^2$$

بمعرفتنا قيمة X واستعمال خط التسوية .

C . تجزئة التباين الهامشي

لقد سمح لنا تحديد منحنى انحدار Y حسب X بتجزئة تباين Y الهامشي إلى مجموع عنصرين : التباين المفسّر بمنحنى الانحدار والتباين المتبقّي حول منحنى الانحدار (أنظر القسم II ، ص 177) .

بطريقة مماثلة ، يمكن تجزئة تباين Y الهامشي بإدخالنا خط تسوية Y حسب X على طريقة المربّعات الصغرى .

بالفعل يمكننا أن نكتب :

$$V(Y) = \frac{1}{n}\sum_i \sum_j n_{ij}(y_j - \bar{y})^2$$
$$= \frac{1}{n}\sum_i \sum_j n_{ij}\{[(y_j - \bar{y}) - a(x_i - \bar{x})] + a(x_i - \bar{x})\}^2$$
$$= \frac{1}{n}\sum_i \sum_j n_{ij}[(y_j - \bar{y}) - a(x_i - \bar{x})]^2 + \frac{1}{n}\sum_i \sum_j n_{ij}[a(x_i - \bar{x})]^2$$
$$+ 2a.\frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})[(y_j - \bar{y}) - a(x_i - \bar{x})].$$

إلاّ أنّ :

$$(y_j - \bar{y}) - a(x_i - \bar{x}) = y_j - ax_i - b$$
$$a(x_i - \bar{x}) = ax_i + b - \bar{y}$$

بناء على تعريف b :

$$b = \bar{y} - a\bar{x}$$

والعبارة :

$$2a.\frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})[(y_j - \bar{y}) - a(x_i - \bar{x})]$$
$$= 2a.\frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})(y_j - \bar{y}) - 2a^2\frac{1}{n}\sum_i \sum_j n_{ij}(x_i - \bar{x})^2$$
$$= 2a\,\mathrm{cov}(XY) - 2a^2\sigma_x^2$$

تساوي صفراً بناء على تعريف a :

$$a = \frac{\mathrm{cov}(XY)}{\sigma_x^2}.$$

نحصل على :

$$V(Y) = \frac{1}{n}\sum_i \sum_j n_{ij}(y_j - ax_i - b)^2 + \frac{1}{n}\sum_i n_{i.}(ax_i + b - \bar{y})^2$$

العبارة الأولى :

$$\frac{1}{n}\sum_i \sum_j n_{ij}(y_j - ax_i - b)^2$$

هي كناية عن حصّة التباين الهامشي الناجمة عن تشتّت النقاط الملحوظة حول خطّ المربّعات الصغرى ، إنّها التباين المتبقّي الذي لا تفسّره العلاقة الخطّية ، وقيمتها هي الحدّ الأدنى المحسوب في الفقرة السابقة :

$$V_R = (1 - r^2)\,V(Y).$$

العبارة الثانية :

$$\frac{1}{n}\sum_{i} n_{i.}(ax_i + b - \bar{y})^2$$

هي كناية عن حصّة التباين الهامشي التي يفسّرها خطّ المربّعات الصغرى لدينا ، بالتالي :

$$V(Y) = (1 - r^2)\ V(Y) + r^2\ V(Y)\ .$$

التباين المفسّر بالخطّ   التباين المتبقّي

التباين الكلّي (التباين الهامشي) يساوي مجموع التباين المفسّر بخطّ المربّعات الصغرى مع التباين المتبقّي .

تُظهِر هذه التجزئة أنّ مربّع معامل الارتباط الخطّي يساوي نسبة تباين Y الهامشي التي يفسّرها خطّ المربّعات الصغرى Y حسب X .

إذا قارنا بين هذه الخاصّة لمعامل الارتباط الخطّي وتعريف نسبة ارتباط Y حسب X (أنظر القسم II ، ص 177) ، يظهر لهذين المؤشّرين المدلول نفسه :

في حالة ارتباط خطّي ، يتطابق خطّ المربّعات الصغرى مع منحنى الانحدار ويكون لدينا $r^2 = \eta^2_{Y/X}$. إذا لم يكن الارتباط خطّياً ، يكون r^2 أصغر من $\eta^2_{Y/X}$ لأنّ التباين المتبقّي يكون حدّاً أدنى بالنسبة لمنحنى الانحدار .

إذا ب دلنا X مع Y ، نحصل على تجزئة تباين X الهامشي بالنسبة لخط المربعات الصغرى X حسب Y :

$$V(X) = (1 - r^2)\ V(X) + r^2\ V(X)$$

$r^2V(X)$ هي التباين المفسّر بخطّ تسوية X حسب Y ، و$(1-r^2)\ V(X)$ التباين المتبقّي حول هذا الخط . إذاً مربّع معامل الارتباط الخطّي يساوي أيضاً لنسبة تباين X الهامشي المفسّر بخط المربّعات الصغرى X حسب Y .

الفصل الخامس

البحث الإحصائي

يمكن القيام بجمع المعلومات حول مجتمع إحصائي معيّن ، إمّا على نحو شامل إمّا على قسم فقط من المجتمع .

إنّ التحقيقات الشاملة ، أو الكشوفات ، تقوم على ملاحظة جميع الوحدات التي تؤلّف المجتمع . وبالطبع ، عندما يكون حجم هذا المجتمع كبيراً ، فإنّ هذه التحقيقات تصبح باهظة الكلفة . ومثل نموذج على هذا الأمر هو الفرز الشامل للجمهور .

أمّا التحقيقات التي لا تتعلّق سوى بقسم من المجتمع الإحصائي ، فلا أهمّية لها إلاّ إذا تمّ اختيار هذا القسم كي يمثّل المجتمع تمثيلاً صادقاً ، بعبارة أخرى كي يمكن بسط المعلومات المجموعة على كلّية المجتمع . ونطلق على هذا النهج اسم الاستقصاء بواسطة البحث الإحصائي .

القسم I

مدخل إلى طريقة البحوثات الإحصائية

1 . حسنات الإستقصاء بواسطة البحث الإحصائي : A . الكلفة والسرعة ؛ B . المرونة في اختيار المفاهيم ؛ C . دقّة وغنى الملاحظات .ـ 2 . حدود الاستقصاء بواسطة البحث الإحصائي : A . أخطاء المعاينة ؛ B . مصاعب اختيار العيّنة .ـ 3 . مختلف أنواع الأبحاث الإحصائية .

البحث أو التحقيق الإحصائي هو بحث يجري على قسم يمثّل المجتمع الإحصائي موضع الدراسة الذي نسمّيه المجتمع المرجع . هذا القسم هو العيّنة . ويسمّى خارج قسمة مقدار العيّنة n على مقدار المجتمع N ، أي t=n/N ، بمعدّل البحث الإحصائي .

وبناءً على تمثيل العيّنة للمجتمع ، تسمح لنا المشاهدات التي نجريها عليها بتقدير توزيع المجتمع المرجع ومقاييسه .

1 . حسنات الإستقصاء بواسطة البحث الإحصائي

البديل عن جمع المعلومات الإحصائية بواسطة البحث الإحصائي هو :

- إمّا القيام بتحقيقات شاملة مناسبة لهذا الغرض تكون شاقّة ومكلِفة ؛
- إمّا استعمال توثيق جُمّع لحاجة الأعمال الإدارية . هكذا ، يجري إحصاء الرواتب في فرنسا انطلاقاً من جداول DAS ، أي بيانات الرواتب المدفوعة موضوعة مع الأحكام الأميرية وأحكام الضمان الإجتماعي من قبل المستخدِمين . في هذه الحالة ، تتعلّق طبيعة ومدلول المعلومات المحصورة بشدّة بالقوانين والأعراف التي تحكم عمليّة استقاء المعلومات .

بالنسبة لهذه الطرق ، تقدّم الأبحاث الإحصائية ميّزات من ناحية الكلفة ، السرعة والليونة . وتسمح من ناحية أخرى بإجراء المشاهدات ، المتعلّقة بعدد وحدات إحصائية صغير نسبياً ، بعناية أكثر وبسطها على عدد أكبر من الخصائص .

A . الكلفة والسرعة

لنفترض أنّ وزارة الإسكان تنوي القيام بدراسة إمكانية توسيع برنامج إعداد أماكن سكنية بثمن رخيص . حتماً سيكون من المفيد لها أن تعرف مسبقاً الاحتياجات (المساحة ، عدد الغرف ، الخ . .) ، الأذواق (منزل مستقلّ ، شقّة ، الخ . .) وإمكانيات الجمهور المادية في ما يخصّ السكن . يمكن النظر في حلّين :

- إمّا القيام بتحقيق شامل عن طريق سؤال كل الأسر ؛
- إمّا اعتماد نهج البحث الإحصائي فلا يُسأل ، مثلاً ، سوى أسرة من كلّ ألفي أسرة .

قد يوجد أكثر من 17 مليون أسرة : يمكننا تصوّر الوسائل المادية والأوقات الضرورية لاعتماد الحلّ الأوّل .

أمّا إذا اعتمدنا طريقة البحث الإحصائي ، يصبح عدد المقابلات التي يجب إجراؤها صغيراً نسبياً : أقل من 9000 . وبواسطة باحث مختصّ ، يتراوح سعر التكلفة الوحدوي لهذا التحقيق من 30 إلى 80F ، تبعاً لتعقيد لائحة الأسئلة . حتّى ولو بدا هذا السعر الوحدوي مرتفعاً ، فإنّ الكلفة العامّة تبقى « معقولة » ، إذا أخذنا بعين الاعتبار أهمّية المعلومات المحصودة ، وعلى أيّ حال لا يمكن قياسها مع كلفة التحقيق الشامل .

والعديد من التحقيقات حول السوق أو استطلاعات الرأي التي تُجرى غالباً على عيّنات صغيرة (2000 أو 3000 وحدة إحصائية) ، تكلّف أقلّ بكثير .

إنّ تحقيقاً دون صعوبة خاصّة ، نجريه على عيّنة صغيرة ، يمكن القيام به بسرعة فيعطي النتائج الأولى خلال مهلة قصيرة : إذ تنجز شركات متخصّصة بعض الدراسات على السوق في غضون أسابيع قليلة ، ويتمّ فرز الاستفتاءات الانتخابية ، المدروسة خصّيصاً لهذه الغاية خلال بضعة أيام .

B . المرونة في اختيار المفاهيم والتصوّرات

نلمس هذه الميزة على نحو ظاهر خاصّة بالنسبة للمعلومات المستقاة من أحد النشاطات الإدارية . في الواقع ، إنّ هذه العمليّات ، عندما لا تحكمها نصوص إلزامية، تنظيمية أو تشريعية ، تخضع على أيّ حال لمجموعة من القواعد : تعريفات ، مصطلحات ، إجراءات تسجيل وفحص ، الخ . . ، لا تكون دائماً ملائمة من وجهة النظر الإحصائية .

من جهة أخرى ، تكون هذه القواعد عرضة للتغيّر مع الوقت والمكان ، من مؤسسة أو من بلد لآخر ، مما يجعل تأويل النتائج صعباً .

هكذا ، منذ نهاية 1967 ، تسبّب تليين شروط قبول العمّال المحرومين من العمل لصالح المساعدة العامّة وتوسيع ضمان البطالة وإنشاء وكالة الاستخدام الوطنية في إخلالات مهمّة على صعيد سلاسل البطالة التي وضعتها دوائر التوظيف الفرنسية الرسمية . هذه التغيّرات التي ليس لها مدلول اقتصادي جعلت خلال سنوات عديدة من الصعب تفسير ظروف هذه الإخلالات . بالمقابل فإن مفهوم البطالة المعتمد في تحقيقات I.N.S.E.E الإحصائية ، المستقلّة عن أيّ مرجع تنظيمي أو مؤسّسي ، لم يتأثّر بهذه التعديلات .

وبنفس هذه التصوّر ، بادر المكتب الإحصائي لدول السوق الأوروبية المشتركة،

للحصول على البيانات الخاصّة بالبلدان الأعضاء ، بإطلاق التحقيقات الإحصائية في مجالات مختلفة ، وذلك بتحديدات متشابهة وطرق متقاربة :

- تحقيقات حول نفقات الأسر ،
- تحقيقات حول الاستخدام ،
- تحقيقات حول كلفة اليد العاملة وحول بنية الأجور ، الخ

C . دقّة وغنى الملاحظات

بحكم حجمه ، يسمح التحقيق الإحصائي باستدعاء باحث مختصّ (تحقيق اجتماعي - اقتصادي ، تحقيق حول السوق) أو جهاز موظّفي ذي نوعية جيّدة (لفحص الصناعات) كما يسمح بالقيام بملاحظة دقيقة ومتزامنة لخصائص علّة .

هكذا ، فإن تحقيقاً حول الاستهلاك يسمح بالحصول ، بالنسبة لكلّ أسرة على :

- خصائصها الاجتماعية - الديموغرافية : عدد الأفراد والأعمار ، الفئة الإجتماعية - المهنية ، المنطقة وفئة مكان السكن ؛
- مدخولها السنوي ؛
- تجهيزها بالمتلكات المستديمة (برّاد (ثلاّجة) ، غسّالة ، سيّارة، جهاز تلفزة ، الخ . . .) مع تاريخ شرائها ،
- نفقاتها المفصّلة على مدّة محدّدة .

وخلال تحقيق حول الركائز الدعائية[1] ، حصلنا بالنسبة لكل وحدة إحصائية من العيّنة على :

- خصائصها الاجتماعية - الديموغرافية العامّة : الجنس ، العمر ، الفئة الإجتماعية - المهنية ، مستوى التعليم ، الاستهلاكات المعتادة ، مكان الإقامة ، الخ . .
- عدد وطبيعة القراءات (بالنسبة لعدد معيّن من الجرائد أو المجلاّت) ؛
- عدد المرّات التي يذهب فيها إلى السينما ؛
- البرامج التي يستمع إليها من الراديو أو التي يشاهدها في التلفزيون .

وتسمح هذه المعلومات أن نحسب مثلاً :

- كم من أفراد فئة معيّنة (تسكن في بلدة ما ، تملك سيارة أو لها عادات استهلاكية

(1) كتاب ج. ديزابل J.Dezable ، « دراسة حول قراءة الصحف » ، مجلّة شركة الإحصاء الباريسية ، تموز - أيلول 1960

معيّنة ، الخ . .) وصل إليه بثّ رسالة دعائية على جهاز معيّن (مثلاً ، الجريدة أ) ؛

- كم من الأفراد هم عرضة لأن تصل إليهم رسالة دعائية مطلقة بواسطة أجهزة مختلفة في آن واحد (مثلاً ، الجريدتين أ وب ، المجلّة ج والتلفزيون) .

ونلمس أهمّية هذا النوع من المعلومات بالنسبة لدراسات العرض والطلب وتنظيم الحملات الانتخابية : تحديد الجمهور - الهدف ، دراسة مقارنة لكلفة وفعالية الأجهزة الإعلامية ، الخ . .

2 . حدود الأبحاث الإحصائية

تتعلّق حدود الأبحاث الاحصائية بشكل أساسي بأخطاء المعاينة ومصاعب تحديد العيّنة .

A . أخطاء المعاينة

ترتكز الأبحاث الإحصائية على قانون الأعداد الكبيرة : لا يمكن تعميم الكمّيات المقاسة على العيّنة إلى المجتمع المرجع وبدقّة مقبولة إلا انطلاقاً من عيّنات ذات حجم كبير بشكل كاف .

إذاً لا يمكن تطبيق طريقة الأبحاث الاحصائية على مجتمعات مقدارها ضعيف : يجب ملاحظتها بشكل شامل . ينبغي أيضاً اتخاذ بعض الاحتياطات عندما يكون المجتمع الإحصائي مؤلّفاً من وحدات غير متساوية الأحجام ، مثلاً مؤسّسات صناعية كثيرة الاختلاف من ناحية الأهمّية . إن طريقة البحوثات الإحصائية تبقى صالحة للتطبيق في هذه الحالة ولكنّها ، كي تكون دقيقة ، تتطلّب معرفة تقريبية لحجم كلّ وحدة بغية أخذه بعين الاعتبار عند سحب العيّنة (أنظر « تفريع العيّنات » ، الفصل VII ، القسم II ، الفقرة 1 ، ص 335) : حيث يجب أخذ معدّل بحث مرتفع أكثر بالنسبة للمؤسّسات الأكثر أهمّية .

من ناحية أخرى ، حتّى حين يكون المجتمع الإحصائي كبيراً ومؤلّفاً من وحدات يمكن مقارنة أحجامها ، لا يمكن تقديم النتائج إلاّ على مستوى معيّن من التجميع : فبحكم أخطاء المعاينة ، قد لا تصبح النتائج المفصّلة كثيراً معبّرة وكاشفة .

B . مصاعب تحديد العيّنة

في بعض الحالات ، قد تصبح طريقة الأبحاث الاحصائية صعبة التطبيق بسبب مصاعب حصر المجتمع المرجع .

لنفترض مثلاً أنّنا نريد إجراء دراسة معمّقة حول البطالة . يتوزع العاطلون عن العمل على مجمل الأقاليم ولا نعرف عناوينهم ، باستثناء المسجّلين عند وكالة الاستخدام الوطنية . يجب إذن الانطلاق من عيّنة كبيرة جدّاً تغطّي كامل المجتمع كي نأخذ منها عيّنة مفيدة ذات حجم كاف . لمؤسّسة صحفية تودّ إجراء استفتاء لقرّائها تصطدم ، إلاّ بالنسبة للمشتركين ، بنفس العوائق . من هنا يكون أحياناً من المفيد أكثر إجراء بعض الاستفتاءات على مستوى المهنة ككل : إنّها حالة الوسائل الدعائية المذكورة أعلاه .

غالباً ما تصادف هذه العوائق في مجال الدراسات حول السوق ، حيث تزيد منها أحياناً عدم دقّة المجتمع المرجع . كي ندرس سوق مادّة جديدة مثلاً ، يجب البدء بتحديد مجموعة الشرّاء المحتملين ، مثلا المؤسّسات التي قد تستعملها في صناعتها . قد يكون من الضروري القيام ببحث تمهيدي للاحاطة بمجال الدراسة ، ثمّ فقط في مرحلة ثانية ، يأتي دور الدراسة الخالصة عن السوق .

والمصاعب تصبح أكبر في ما يخصّ الأبحاث الإحصائية العشوائية : حيث يجب أن يكون بمتناولنا قاعدة للبحث العشوائي ، أي لائحة أو ملفّ يسمح بمعاينة الوحدات المنتمية إلى المجتمع المرجع دون حذف ودون تكرار .

3 . مختلف أنواع الأبحاث الإحصائية

يمكن التمييز بين فئتين كبيرتين من الأبحاث الإحصائية : الأبحاث على أساس « الاختيار المدروس » والأبحاث « العشوائية » .

الأبحاث على أساس مدروس تعني مختلف التقنيات التي تقوم على بناء ، انطلاقاً من معلومات مسبقة حول المجتمع الإحصائي موضع الدراسة ، عيّنة شبيهة قدر الإمكان بهذا المجتمع . يأتي تحديد العيّنة نتيجة اختيار مدروس ومن هنا اسم الطريقة . إنّها مناهج تجريبية تتضمّن قسماً من الاعتباطية ولا تسمح بتقييم دقّة التقديرات . إلاّ أنّها لها حسناتها ، خاصّة من ناحية الكلفة والسرعة ، بالمقارنة مع طريقة الأبحاث العشوائية .

الأبحاث العشوائية هي مجموعة طرق سحب العيّنة حيث كلّ من وحدات المجتمع الإحصائي لها احتمال معروف ، مختلف عن الصفر ، لأن تنتمي إلى هذه العيّنة . المتغيّرات الملحوظة على العيّنة هي متغيّرات عشوائية : بناءً على هذه المتغيّرات ، لا يمكن تقدير الكمّيات المناسبة المتعلّقة بمجمل المجتمع الإحصائي

وحسب ، بل أيضاً أن ننسب لهذه التقديرات قياساً للخطأ الممكن ارتكابه .

القسم II

طريقة الكوتا (أو الأنصبة)

1 . مبدأ طريقة الكوتا .ـ 2 . تطبيق الطريقة : A . اختيار متغيّرات المراقبة ؛ B . تنظيم البحث عملياً ؛ C . مراقبة الباحثين .ـ 3 . حسنات وسيّئات طريقة الكوتا : A . الحسنات ؛ B . السيئات .

تقوم الطرق التجريبية لتحديد العيّنة باستدعاء « الاختيار المدروس » : نختار العيّنة بشكل يؤلّف صورة ، صادقة قدر الإمكان ، عن المجتمع الإحصائي . والتقنية التي يكثر من استعمالها عادة هي طريقة الكوتا .

1 . مبدأ طريقة الكوتا

تفترض طريقة الكوتا ، المستعملة عادة في الدراسات الاجتماعية ـ الاقتصادية (دراسات حول السوق ، استفتاءات الآراء ، الخ . .) وجود ارتباط بين مختلف خصائص المجتمع الإحصائي . إذا ثبتت صحّة هذا الافتراض ، فإنّ عيّنة مأخوذة بشكل تمثل فيه توزيعاً إحصائياً لبعض الخصائص المختارة عن سابق تصوّر شبيهاً بتوزيع المجتمع الإحصائي ، لها أيضاً فرص كبيرة بأن تكون قريبة جدّاً من هذا المجتمع في ما يتعلق بتوزيع خصائص أخرى .

إنّ الخصائص التي نأخذها لتأمين مشابهة العيّنة لمجمل المجتمع الإحصائي نسمّيها متغيّرات المراقبة أو متغيّرات الفحص .

وكي نكون قادرين على تطبيق طريقة الكوتا ، يجب معرفة توزيع المجتمع الإحصائي حسب متغيّرات المراقبة . ونحصل على الكوتا ، التي يجب أن يراعيها الباحثون ، بضربنا مقادير مختلف كيفيّات متغيّرات المراقبة بمعدّل البحث الإحصائي . بهذه الطريقة نضمن للعيّنة نفس بنية المجتمع الإحصائي من ناحية متغيّرات المراقبة . وضمن إطار الكوتا يُترك أمر اختيار أفراد أو وحدات العيّنة لتقدير الباحث .

مثلاً لنفترض أنّ مجمعاً متخصّصاً كُلّف بدراسة انتشار صحيفة يومية محلّية بين سكان منطقة تولوز (Toulouse) . متغيّرات المراقبة المختارة هي الجنس ، العمر والفئة الإجتماعية المهنية ، ومعدّل البحث الإحصائي المأخوذ هو t= 1/300 ، بشكل يكون فيه

مقدار العيّنة قريباً من الألف .

يعطينا إحصاء 1968 توزيعات السكّان البالغة أعمارهم أكثر من 15 سنة في هذه المنطقة حسب متغيّرات المراقبة (الجدول 23) . إذا ضربنا المقادير المناسبة بمعدّل البحث الإحصائي ، نحصل على الكوتا المعدّة لتأمين التشابه ، من ناحية متغيّرات المراقبة ، بين بنية العيّنة وبنية المجتمع الإحصائي (الجدول 24 ، العواميد (1)) : نستجوب ما مجموعه 1154 شخصاً ، يجب أن يتضمّنوا 544 رجلاً ، 195 شخصاً تتروح أعمارهم بين 25 و34 سنة ، 200 عامل، الخ . . تُملى هذه الكوتا إذن على الباحثين : يحصل كلّ واحد منهم على جدول مراقبة يشير عليه كم شخصاً من كلّ فئة يجب أن يستجوب . هكذا ، نسلّم إلى باحث عليه إجراء 50 مقابلة جدول مراقبة يطابق العواميد (2) من الجدول 24 .

2 . تطبيق الطريقة

A . اختيار متغيّرات المراقبة

كي يمكننا أخذ خاصّة إحصائية معيّنة كمتغيّرة مراقبة ، عليها أن تملأ الشروط التالية :

- أن تكون على ارتباط وثيق بالمتغيّرات موضع الدراسة ؛
- أن يكون توزيعها الإحصائي على مجمل المجتمع معروفاً ؛
- أن تنسجم مع ملاحظة الباحثين على أرض الدراسة دون احتمالات خطأ مفرِطة .

إنّ المبدأ الأوّل يعبّر عن شرط فعالية الطريقة نفسه ، ويوضّح المبدآن الآخران شروط إمكانية تطبيقها . المدخول ، مثلاً ، لا يمثل بشكل عام متغيّرة جيّدة للمراقبة ، في الواقع حتّى ولو كان هذا المقياس ممتازاً بالنسبة للشرط الأوّل ، خاصّة في ما يتعلّق بدراسات السوق (العرض والطلب) ، فإنّ توزيعه غير معروف كلّياً وملاحظته من قبل الباحث صعبة . لهذا السبب نفضّل بشكل عام استبداله بالفئة الاجتماعية ـ المهنية . يجب أيضاً أن يتمّ تحديد فئة فرد معيّن على أساس قواعد دقيقة ، مطابقة لّلتي استعملتها المؤسّسة الإحصائية والّتي وجدنا الكوتا بواسطتها . فإنّ أخطاء التصنيف قد تسبّب بخطأ منهجي[1] في النتائج .

---

(1) في المثل السابق يجب على الباحث أن يستعمل ، لتصنيف فرد ما ضمن فئة اجتماعية ـ مهنية معيّنة ، نفس القواعد المستعملة في فرز السكّان العام . إذا كان الباحث يميل إلى وضع ، في فئة « العمّال » ، أشخاص صنّفوا « موظّفين » في الفرز العام ، ينتج عن هذا تغيّر في صورة العيّنة : إذ يكون تمثيل العمّال (في الفرز العام) ناقصاً وتمثيل الموظّفين زائداً . بالتالي قد يشوب النتائج خطأ منهجي .

الجدول 23 . توزيع سكّان منطقة تولوز ، من 15 سنة وأكثر ، حسب الجنس العمر والفئة الإجتماعية ـ المهنية .

المصدر : كشف I.N.S.E.E للسكّان 1968
الوحدة : ألف

الجنس			العمر			الفئة الإجتماعية ـ المهنية		
		%			%			%
مذكّر	163,2	47,1	15 إلى 24 سنة	81,6	23,6	أرباب عمل [0+2]	19,2	5,5
مؤنّث	183,2	52,9	25 إلى 34 سنة	58,5	16,9	مهن حرّة وكوادر عليا [3]	14,6	4,2
			35 إلى 54 سنة	107,4	31,0	كوادر وسط وموظّفون [4+5+7+8]	78,1	22,6
			55 سنة وأكثر	98,9	28,5	عمّال [1+6]	60,2	17,4
						أصحاب دخل، متقاعدون عاطلون عن العمل [9]	174,3	50,3
المجموع	346,4	100,0	المجموع	346,4	100,0	المجموع	346,4	00,0

الجدول 24 . الكوتا العائدة لمنطقة تولوز بالنسبة لمجمل العيّنة (معدّل البحـ t= 1/300) ولـ 50 مقابلة .

الجنس			العمر			الفئة الاجتماعية ـ المهنية		
	(1)	(2)		(1)	(2)		(1)	(2)
مذكّر	544	24	15 إلى 24 سنة	272	12	أرباب عمل [0+2]	64	3
مؤنّث	610	26	25 إلى 34 سنة	195	8	مهن حرّة وكوادر عليا [3]	49	2
			35 إلى 54 سنة	358	16	كوادر وسط وموظّفون [4+5+7+8]	260	11
			55 سنة وأكثر	329	14	عمّال [1+6]	200	9
						أصحاب دخل ، متقاعدون عاطلون عن العمل [9]	581	25
المجموع	1154	50	المجموع	1154	50	المجموع	1154	50

إنّ هذه الشروط تحدّ كثيراً من حرّية اختيار متغيّرات المراقبة ، ومن المتغيّرات المستعملة دوماً يمكننا أن نذكر :

- بالنسبة لعيّنة من الأشخاص : الجنس ، العمر ، الفئة الاجتماعية - المهنية ، المنطقة ، فئة المنطقة (مناطق مدينية أم ريفية) ؛
- بالنسبة لعيّنة من الأسر : فئة ربّ العائلة الاجتماعية - المهنية ، عدد أعضاء الأسرة ، المنطقة ، فئة المنطقة ؛
- بالنسبة لعيّنة من نقاط المبيع : نوع التجارة (حرّة أم غير حرّة) ، عدد الأجراء ، طبيعة النشاط التجاري ، المنطقة ، فئة المنطقة .

بالطبع ، بناء على المبدأ الأوّل ، يجب أن يتمّ اختيار متغيّرات المراقبة تبعاً لموضوع الدراسة : مثلاً ، بالنسبة لبحث حول نفقات السكن ، قد يكون من المهمّ مراقبة عدد الأسرة المستأجرة لمسكن جديد ، لمسكن قديم ، الأسر المالكة ، الخ . .

B . تنظيم البحث عملياً

أ - تحديد العيّنة : بحث على عدّة درجات

غالباً ، لا يكون مجال الدراسة عبارة عن تجمّع واحد (تولوز) ، بل بلد بأكمله ، فرنسا مثلاً ، أو منطقة بأكملها (الجنوب والبيرنيه ، Midi-Pyrénées) ، ويتضمّن عدداً كبيراً من النواحي . من غير المعقول طبعاً إجراء البحث في كلّ من هذه النواحي : إذ تصبح نفقات التنقّل مرتفعة جداً .

عملياً ، نعمد عامّة إلى بحث بدرجتين : نبدأ عند درجة البحث الأولى بتحديد عيّنة من النواحي (وحدات أوّلية) ، ثمّ ، ضمن النواحي - العيّنة ، نختار عند الدرجة الثانية من البحث عيّنة من الوحدات الثانوية : أشخاص ، أُسَر ، نقاط مبيع ، مؤسّسات صناعية ، . . . حسب طبيعة الحملة .

إنّ اختيار النواحي - العيّنة هو على أهمّية كبيرة ، ونجريه باستعمالنا عدد معيّن من متغيّرات المراقبة ينتج عن تلاقيها فروع . نعتمد بشكل عام متغيّرات المراقبة التالية : - المنطقة : يمكننا مثلاً تقسيم فرنسا لهذا الهدف إلى 8 مناطق كبيرة ؛

- فئة المنطقة . يمكننا مثلاً التمييز بين :

● المناطق الريفية (حيث تجمّع السكّان في مركز القضاء يعدّ أقلّ من 2 000 نسمة) ؛

- المدن الصغيرة : من 2000 إلى 10 000 نسمة ؛
- المدن أو التجمّعات من 10 إلى 20 000 نسمة ؛
- المدن أو التجمّعات أكثر من 50 000 نسمة .

بهذه الطريقة نحدّد بالنسبة لفرنسا بكاملها (32 = 4×8) 32 فرعاً نعيّن ضمنها النواحي ـ العيّنة . ويمكننا ، بطبيعة الحال ، إدخال كلّ من التجمّعات التي تعدّ أكثر من 50 000 نسمة ضمن العيّنة ، وبالمقابل لا نحتفظ في هذه العيّنة إلاّ بجزء من المدن أو النواحي التي تنتمي إلى الفروع الأخرى .

ب ـ كيفيّات تنظيم البحث

إنّ تنظيم البحث يتعلّق كثيراً بتكوين شبكة الباحثين .

ـ يمكننا استعمال شبكة دائمة من الباحثين يعملون في محيط سكنهم ، ويسمح هذا الإجراء بتنقيص سعر تكلفة الحملات عن طريق تخفيض نفقات التنقّل . وتكون عيّنة الوحدات الأوّلية (عيّنة النواحي) مشتركة بين كلّ الحملات وتمثّل العيّنة ـ الرئيسة . ويتمّ وضع شبكة الباحثين نهائياً تبعاً لهذه العيّنة ـ الرئيسة من النواحي .

حسب طريقة التنظيم هذه ، لا يعمل كلّ باحث سوى في ناحية واحدة ، يجب إذاً وضع الكوتا كلاًّ على حدة لكلّ من هذه النواحي .

ـ يمكننا بالمقابل استعمال فرق من الباحثين المتنقّلين ، يديرها المشرف أو رئيس البحث ، وتغطّي كلّ منها قسماً واسعاً من المكان الخاضع للدراسة . إنّ هذه الطريقة مكلفة أكثر لأنّ نفقات النقل تكون مرتفعة جدّاً ، ولكنّها أكثر مرونة . يمكننا بصورة خاصّة وضع كوتا لمنطقة بأكملها .

لنأخذ مثل حملة تغطّي منطقة الجنوب والبيرينيه . بالإضافة إلى تولوز يوجد في هذه المنطقة تجمّعان آخران يعدّان أكثر من 50 000 نسمة ، تارب (Tarbes) وألبي (Albi) اللذان نأخذهما بأكملهما ضمن العيّنة . ونحدّد في الفروع الأخرى النواحي ـ العيّنة .

سنملي ، من جهة ، على فريق الباحثين توزيع الحملات بين النواحي :

1 154 مقابلة في تولوز
187 في تارب
140 في ألبي
إلخ . . . ،

ومن جهة أخرى الكوتا حسب الجنس ، العمر والفئة الاجتماعية المهنية ، التي وضعناها لمجمل المنطقة .

C . مراقبة الباحثين

خلال حملة تتبّع البحث العشوائي ، يعمل الباحثون على أساس لوائح لعناوين الأشخاص أو الوحدات التي يجب إجراء الدراسة عليها ومن السهل التحقّق ما إذا كانوا يلتزمون بهذه اللوائح . أمّا في حملة تتبّع طريقة الكوتا من الصعب مراقبة الطريقة التي يختار بها الباحث الأشخاص الذين يستجوبهم وبشكل خاص ما إذا كان يتقيّد بالكوتا : ويكون من الفطنة أن نطلب من الباحثين أن يدوّنوا اسم وعنوان الأشخاص المستجوبين بشكل يؤمّن لنا إمكانية المراقبة . على كلّ حال ، أن نترك للباحثين المبادرة في اختيار وحدات العيّنة هو أمر يزيد من قابلية التغيّر بشكل ملحوظ .

فكّرنا إذاً بالحدّ من الحرية المتروكة للباحثين وذلك كي نقلّل من تأثيرها على النتائج .

من الجيّد مثلاً أن نملي على الباحثين ، عدا عن ضرورة التقيّد بالكوتا ، عدداً من الشروط الإضافية :

- منع انتقاء الأشخاص الذين سيُستجوبون تبعاً للوائح معيّنة : لوائح المشتركين ، الزبائن ، الأشخاص الذين طلبوا سلعة معيّنة إلى منزلهم ، . . . إذ يوجد بين هؤلاء الأشخاص في الواقع شيء مشترك : فهم يقرؤون جريدة كذا أو اشتروا مؤخّراً برّاداً معيّناً . ويمكننا تصوّر سيّئات هذه اللوائح ، حتّى ولو اتُّبعت الكوتا بكلّ دقّة ، إذا كان موضوع الحملة على علاقة مع المبدأ الذي وضعت على أساسه : مثلاً انتقاء الأسر المستجوبة لدراسة حول نسبة امتلاك هاتف وذلك في دليل الهاتف ؛
- منع العمل في الشارع : من أجل دراسة حول وسائل التسلية ، يمكن للباحث أن يتقيّد جيّداً بالكوتا ويكتفي باستجواب الأشخاص المنتظرين على أبواب صالات السينما !
- منع إعادة استجواب نفس الأشخاص .

غالباً ما يُعمد بالنسبة للحملات المدينية إلى نهج يحدّ من حرّية الباحثين في اختيار الأسر التي ستُستجوَب وهو طريقة بوليتز (Politz) ، التي تملي على كلّ باحث خطّ سير يُحدَّد بدقّة ويدلّه على نقاط البحث .

من وجهة نظر الباحث يجري الأمر كما لو كانت العيّنة عشوائية : نملي عليه لائحة

من المساكن التي سيزورها وذلك بعد أن نعاينها بواسطة إحداثياتها الجغرافية . بالتالي يمكننا مراقبة عمله .

في الحقيقة العيّنة ، طبعاً ، ليست عشوائية لأنّه ليس لكلّ المساكن نفس الاحتمال لان نأخذها . إذاً يتوقّف حسن تمثيل العيّنة فقط على مهارة من يضع خطّة البحث الإحصائي .

بعكس طريقة الأبحاث الإحصائية العشوائية ، فإنّ هذه الطريقة لا تستدعي وجود قاعدة للبحث . ورغم كونها أكثر كلفة من مجرّد طريقة الكوتا فهي تبدو أكيدة أكثر وتُستعمل أكثر فأكثر من قبل الأجهزة المختصّة بدراسات السوق .

3 . حسنات وسيئات طريقة الكوتا

A . الحسنات

- بخلاف الأبحاث العشوائية ، لا تتطلّب وجود قاعدة بحث ، وهذه ميزة حاسمة كلّياً في حالات عديدة حيث لا وجود لقاعدة بحث أو حيث لا يمكن للجهاز المكلّف بإجراء الحملة أن يستعملها لأسباب تتعلّق بالسّرية الإحصائية .

- إنّ كلفة الأبحاث على طريقة الكوتا هي حتماً أقلّ بكثير من كلفة الأبحاث الاحتمالية . فبحكم تخفيض التنقّلات يكون مردود الباحث مضاعفاً تقريباً عندما يترك أمر اختيار الوحدات المستجوَبة لتقديره ولا يكون مفروضاً بواسطة لائحة عناوين .

ونميل في بعض الحالات ، عندما يمكن لأخطاء الملاحظة ، بحكم طبيعة الدراسة ، أن تكون مرتفعة أكثر من أخطاء المعاينة ، إلى اعتماد بحث بواسطة الكوتا بدلاً من بحث عشوائي مكلف أكثر .

B . السيئات

- ليست لطريقة الكوتا أسس نظرية كافية ، فهي تستند على الإفتراض التالي : إنّ التوزيع الصحيح للخصائص المراقَبة يضمن تمثيلاً صحيحاً لتوزيع الخصائص المدروسة . ولكن يمكننا دوماً دحض هذا الافتراض ، وقد شاهدنا أمثلة كاريكاتورية بعض الشيء : انتقاء الأشخاص المستجوَبين من الدليل ، من صفوف الانتظار أمام صالات السينما . . . ولقد أظهرت بعض الدراسات الاختبارية أنّه في غياب فحص دقيق لهذه النقاط ، تميل طريقة الكوتا إلى سوء تمثيل عمّال المصانع وطبقات المجتمع الأقلّ تعلّماً والأشخاص الذين لا يمارسون سوى القليل من النشاطات الاجتماعية ،

الخ . .[1] . بشكل عام ، يميل الباحثون إلى استجواب الأشخاص القريبين من محيطهم الاجتماعي .

. ويكون من الفطنة أن نتحقّق استدلالياً من توزيع متغيّرة أو متغيّرات عدّة غير مراقبة يكون توزيعها من جهة أخرى في المجتمع المرجع معروفاً . ويتكوّن لدينا بهذه الطريقة تخمين ، وليس إثبات ، في ما يخصّ صدق تمثيل العيّنة للمجتمع .

- لا تسمح طريقة الكوتا بحساب دقّة التقديرات التي نحصل عليها انطلاقاً من العيّنة .

بما أنّ الباحثين هم من اختار الأشخاص المستجوبين ، ليس من الممكن معرفة الاحتمال الذي يملكه كلّ فرد من المجتمع الإحصائي في أن ينتمي إلى العيّنة . لا يمكننا إذاً تطبيق حساب الاحتمالات الذي يسمح لنا ، في حالة الأبحاث العشوائية ، بأن نعطي كلّ تقدير قياساً للخطأ الذي قد يُرتكب .

بالخلاصة ، تبدو طريقة الكوتا طريقة تجريبية يمكنها ، رغم افتقارها إلى الأسس النظرية الكافية ، تقديم خدمات قيّمة .

وقد جاء في أحد تقارير اللجنة الإحصائية للأمم المتحدة حول هذا الموضوع :

« يمكن لطريقة الكوتا المستعملة بمهارة أن تعطي فكرة عن أفضليات الجمهور وتغييرات الآراء ، في الحملات البسيطة وعندما لا يكون ضرورة لوجود دقّة عالية . ولكن ليس من الممكن تقييم الدقّة الحاصلة ، ويجدر النظر إلى نتائج البحث بواسطة الكوتا على أنّها ذاتية ، ولا يجب الوثوق بها عندما نكون بحاجة إلى معلومات موضوعية خالية من عوامل الأخطاء الثابتة » .

في غياب قاعدة بحث مناسبة ، هذه الطريقة هي الوحيدة القابلة للاستعمال ، وهي متكيّفة بصورة خاصّة مع الحصول السريع على النتائج مع تقريب كبير ، خصوصاً عندما لا يمكننا بأيّ حال مراقبة الخصائص المدروسة ، بحكم طبيعتها ، بشكل دقيق .

من جهة أخرى وبما أنّ الطريقة العشوائية تستند إلى قانون الأعداد الكبيرة ، عندما يكون مقدار العيّنة صغيراً ، فإنّ خطأ المعاينة يمكن أن يكون أقلّ مع نظام اختيار

---

(1) كتاب ج . ديزابي J. Desabie حول نظرية الأبحاث وتطبيقها ، Dunod ، 1971 .

مدروس منه على السحب العشوائي[1] .

عملياً ولاعتبارات تتعلّق بسعر الكلفة ، فإنّ الوكالات المتخصّصة في الحملات الإحصائية حول آراء الجمهور ودراسات السوق لا تستعمل تقريباً سوى طريقة الكوتا . ولا يمكن إغفال هذه الطريقة مطلقاً بالنسبة للأبحاث ذات الطابع الاجتماعي أو الاقتصادي ، خاصّة عندما نعتقد أنّ بين الأشخاص المسحوبين بالصدفة هناك من سيتهرّب من الاستجواب . هذه مثلاً حالة الأبحاث حول نفقات العائلات ؛ حيث رفض الإجابة يستدعي استبدالات تكون عادةً صعبة المعالجة وتتسبّب بفقدان جزء من حسنات الاختيار بالصدفة .

القسم III

طريقة الأبحاث الإحصائية العشوائية

1 . تعريف .ـ 2 . أساس الطريقة : A . لا مساواة بيانيميه ـ تشيبتشيف B . قانون الأعداد الكبيرة .

1 . تعريف

تتميّز طريقة الأبحاث العشوائية بفعل اختيار العيّنة بشكل يكون فيه لكل وحدة من المجتمع الإحصائي احتمال معروف ، مختلف عن الصفر ، لأن تؤخذ .

عادة ، على الصعيد العملي نخصّص لكلّ وحدة من المجتمع الإحصائي نفس الاحتمال لأن تنتمي إلى العيّنة : يمكننا إذن تشبيه اختيار هذه العيّنة بسحب كرات من وعاء معيّن .

يمكن إجراء السحوبات بطريقتين مختلفتين :

1 . مع ردّ إلى الوعاء (سحوبات مستقلّة أو برنولية)

بعد كلّ سحب نردّ الوحدة التي أخذناها لتوّنا إلى الوعاء قبل أن نعمد إلى اختيار الوحدة التالية . يبقى تكوين الوعاء كما هو ويمكن تعيين كل وحدة من المجتمع المرجع

(1) وذلك عند غياب تفريع عيّنات المجتمع المرجع قبل سحب العيّنة بالقرعة . وإدخال التفريع حسب متغيّرات المراقبة يعود ويرجّح كفّة الطريقة العشوائية .

عدّة مرّات بالقرعة . عدد وحدات العيّنة X التي تمثّل خاصّة معيّنة A هو متغيّرة عشوائية ذات حدّين (أنظر الفصل II ، القسم I)

2 . بدون ردّ إلى الوعاء (سحوبات مستنفِدة)

لا نعيد الوحدة التي سحبناها إلى الوعاء الذي يتغيّر تكوينه بهذه الطريقة عند كلّ سحب . لا يمكن اختيار كلّ وحدة من المجتمع الإحصائي سوى مرّة واحدة وتكون العيّنة مؤلّفة من n وحدة مختلفة يمكننا تعيينها ، بالتالي ، دفعة واحدة . عدد وحدات العيّنة X التي تمثّل خاصّة معيّنة A هي متغيّرة عشوائية فوق هندسية (أنظر الفصل II ، القسم II) .

2 . أساس الطريقة : قانون الأعداد الكبيرة

خلال الفصل I حيث أُدخِلت فكرة توزيع الاحتمال ، قد يكون القارىء لاحظ دون شك صلة القرابة الموجودة بين التصوّرات الإحصائية والتصوّرات الاحتمالية . حيث تتناسب فكرة التردّد أو التكرار بالنسبة للتوزيع الإحصائي الملحوظ مع فكرة الاحتمال بالنسبة لقانون الاحتمال ؛ وتتناسب فكرة وسط المتغيّرة الإحصائية الحسابي مع فكرة أمل المتغيّرة العشوائية الرياضي ، الخ . .

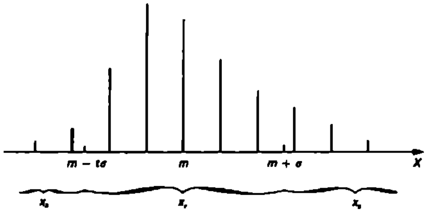
من جهة أخرى وأمام استحالة تحديد ، خاصة في مجال الدراسات التي تهمّ الأعمال ، نظام حوادث متعادلة الاحتمال يمكن حساب احتمالاتها مسبقاً (كما في حالة ألعاب الصدفة) ، اتّجهنا إلى إنشاء نظرية مبدئية لحساب الاحتمالات : الاحتمال المنسوب إلى حدث معيّن هو عدد يخضع لعدد من الشروط أو المبادىء . لكنّ هذه النظرية ليست كافية بحدّ ذاتها لإعطائنا القيمة العددية لاحتمال هذا الحدث ، وحدها المعطيات الملحوظة تسمح بتقدير هذه القيمة . يبقى إذن أن نمدّ جسراً بين المعطيات التجريبية والتصوّرات المجرّدة كي نبرّر هذا الإجراء : إنّه قانون الأعداد الكبيرة الذي قدّمه جاك برنولي (Jacques Bernoulli) منذ بداية القرن XVIII .

A . لا مساواة بيانيميه ـ تشيبيشيف (Bienaymé-Tchebicheff)

لنأخذ متغيّرة عشوائية X ، أملها الرياضي m وانحرافها النموذجي σ لندرس الاحتمال P لأن تنتمي X إلى الفسحة $(m - t\sigma, m + t\sigma)$ المتماثلة بالنسبة للمتوسط :

$$P = P\{\,|X - m|\,\} \leqslant t\sigma ;$$

حيث m و σ هما معطيّان وt عدد يحدّد طول الفسحة (الشكل 53) .



الشكل 53 . لا مساواة بيانيميه ـ تشيبيشيف

كي نختصر من الرموز ، سنستدلّ على P في الحالة حيث المتغيِّرة X منفصلة ، لكن البرهان يبقى صالحاً بالنسبة لمتغيِّرة متواصلة .

بناء على التعريف :

$$\sigma^2 = \sum_i p_i(x_i - m)^2 .$$

لنميِّز بين قيم X الموجودة داخل الفسحة m ± t σ والتي سنرمز إليها بواسطة x_r وبين قيمها الموجودة في الخارج x_s :

$$\sigma^2 = \sum_r p_r(x_r - m)^2 + \sum_s p_s(x_s - m)^2 \qquad (1)$$

إذاً :

$$\sigma^2 \geqslant \sum_s p_s(x_s - m)^2 , \qquad (2)$$

وذلك لأنّ $\sum_r p_r(x_r - m)^2$ هو عدد إيجابي أو يساوي الصفر .

من جهة أخرى الفروقات $x_s - m$ هي ، بحكم تعريفها ، بالقيمة المطلقة أكبر أو تساوي t σ :

$|x_s - m| \geqslant t\sigma$ مهما تكن s .

إذا استبدلنا في (2) $(x_s - m)^2$ بـ $t^2 \sigma^2$ يصبح لدينا من باب أولى :

$$\sigma^2 \geqslant t^2 \sigma^2 \sum_s p_s$$

أي ، إذا قسمنا العنصرين على σ^2 :

$$1 \geqslant t^2 \sum_s p_s$$

$$\sum_s p_s \leqslant 1/t^2. \qquad (3)$$

المجموع $\sum_s p_s$ يمثّل احتمال أن تأخذ X قيمة لا تنتمي إلى الفسحة $m \pm t\sigma$

$$\sum_s p_s = 1 - P.$$

بالتالي إذا انتقلنا إلى (3) :

$$P \geqslant 1 - 1/t^2$$

$$P\{|X - m| \leqslant t\sigma\} \geqslant 1 - 1/t^2.$$

هذه المباينة (لا مساواة) هي مباينة بيانيميه ـ تشيبتشيف ، ومدلولها هو الآتي : إذا كنا نعرف قيمة الانحراف النموذجي σ لمتغيّرة عشوائية معيّنة ، يمكننا دوماً اختيار t كبيرة بشكل كاف كي يكون الاحتمال المنسوب إلى الفسحة $m \pm t\sigma$ ، ومهما يكن قانون احتمال المتغيّرة X موضع الدراسة ، قريباً من 1 قدر ما نريد . بعبارة أخرى ، نكون شبه متأكدين أنّ X تنتمي إلى الفسحة المحدّدة بهذا الشكل . وسيسمح لنا عدم المساواة هذا أن نبرهن قانون الأعداد الكبيرة .

B . قانون الأعداد الكبيرة

أ ـ ميل التردّد الملحوظ لحدث معيّن نحو احتماله

لنأخذ سحب عيّنة مقدارها n من مجتمع إحصائي يتضمّن وحدات A بنسبة p ووحدات B بنسبة $q=1-p$. إذا أجرينا السحب مع ردّ ، فإنّ أمل التردّد $f_n=X/n$ للوحدات A الملحوظة في العيّنة الرياضي هو p وانحرافه النموذجي

$$\sigma = \sqrt{pq/n} \ (^1).$$

لنطبّق مباينة بيانيميه ـ تشيبتشيف :

$$P\{|f_n - p| \leqslant t\sqrt{pq/n}\} \geqslant 1 - 1/t^2. \qquad (4)$$

---

(1) يمكن كذلك إجراء البرهان في حالة عيّنة مسحوبة دون ردّ . عندها يكون انحراف التردّد النموذجي :

$$\sigma = \sqrt{pq/n}\sqrt{(N-n)/(N-1)}.$$

بالتالي :

ـ يمكننا دوماً اختيار t كبيرة بشكل كاف لأن يجعل احتمال أن تنتمي f_n إلى الفسحة $p \pm t\sqrt{pq/n}$ قريباً من 1 قدر ما نرغب .

ـ بعد تثبيت قيمة t ، يمكننا دوماً اختيار مقدار العيّنة n كبيراً بشكل كاف لأن يجعل f_n قريبة من p قدر ما نرغب .

مثلاً . يتضمّن مجتمع إحصائي معيّن نسبة p=0,4 من العناصر A . نرغب في أن ينتمي التردّد f_n للعناصر A الملحوظة في العيّنة إلى الفسحة p±0,01 باحتمال يساوي 99% على الأقلّ :

$$P\{ |f_n - p| \leq 0{,}01 \} \geq 0{,}99$$

لنقارب هذه العبارة مع عدم المساواة (4) :

ـ يجب اختيار t = 10 كي يكون $1 - 1/t^2 = 0{,}99$

ـ بعد تثبيت قيمة t وكي نحصل على : $10\sqrt{pq/n} \leq 0{,}01$ ، $t\sqrt{pq/n} \leq 0{,}01$ ،

يكفي أن نأخذ : $n \geq 240\ 000$

وهذا ما نسمّيه قانون الأعداد الكبيرة : يكفي أن نسحب عيّنة بمقدار كاف من مجتمع إحصائي مركّب على نحو معيّن (يتضمّن نسبة p من الوحدات الإحصائية A) كي يكون تردّد الوحدات A الملحوظ f_n شبه مؤكّد قريباً جدّاً من الاحتمال p .

إلّا أنّه لا يمكن التأكّد مطلقاً من أنّ f_n يوجد في الفسحة المرغوب فيها حول p : واحتمال عدم تحقّق هذا الأمر يساوي $1/t^2$ على الأكثر . ونقول أنّ التردّد الملحوظ لحدث معيّن يميل بالاحتمال نحو احتمال هذا الحدث ، عندما تتزايد n بشكل غير متناه .

إنّ الفائدة الرئيسية من قانون الأعداد الكبيرة هي : إذا كنّا نجهل قيمة الاحتمال p (نسبة الوحدات A في المجتمع الإحصائي) ، يمكننا دوماً أن نأخذ عيّنة عشوائية بمقدار كاف كي يعطي التردّد (التكرار) الملحوظ تقديراً لهذا الاحتمال على قدر ما نريد من الدقّة . هكذا يسمح لنا قانون الأعداد الكبيرة أن نمدّ جسراً بين الصياغة المبدئية لحساب الاحتمالات والتطبيق ، وذلك بإعطائنا وسيلة لنسب قيم عددية لاحتمالات الحوادث موضع الدراسة .

ب ـ ميل الوسط الملحوظ لمتغيّرة عشوائية نحو أملها الرياضي

لنفترض X_1 ، X_2 ، . . . ، X_n ، n متغيّرة مستقلّة تتبع قانون احتمال أمله

الرياضي m وانحرافه النموذجي σ . إنّ متوسّط هذه المتغيّرات :

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

هو نفسه متغيّرة عشوائية أملها الرياضي m وانحرافها النموذجي $\sigma/\sqrt{n}$ أنظر الفصل I ، ص 58 و62) .

لنطبّق عدم مساواة بيانيميه ـ تشيتشيف على هذه المتغيّرة :

$$P\left\{|\overline{X} - m| \leqslant t\frac{\sigma}{\sqrt{n}}\right\} \geqslant 1 - \frac{1}{t^2}.$$

يكفي إذن أن نسحب من المجتمع المرجع عيّنة كبيرة بشكل كاف كي يكون متوسّط المتغيّرة الملحوظ على العيّنة قريباً جدّاً بشكل شبه مؤكّد (مع احتمال يساوي $1-1/t^2$ على الأقلّ) من أملها الرياضي (أي من متوسّط المجتمع الحقيقي) .

هذا النصّ الجديد لقانون الأعداد الكبيرة هو أكثر عموميّة من سابقه . في لواقع ، يمكننا دوماً اعتبار متغيّرة ذات حدّين X كمجموع n متغيّرة برنولي عشوائية (أنظر الفصل II ، القسم I ، الفقرة 3 ، ص 72) وبالتالي اعتبار تردّدها $f_n = X/n$ كمتوسّط هذه المتغيّرات الـ n . يعبّر إذن قانون الأعداد الكبيرة عن ميل متوسّط عيّنة من n مشاهدة ، مأخوذة من مجتمع إحصائي يخضع لقانون احتمال معيّن ، بالاحتمال نحو أمل هذا القانون الرياضي ، عندما تتزايد n بصورة غير متناهية . •

على الصعيد العملي ، يعلّمنا قانون الأعداد الكبيرة أنّه ضمن شرط أن يكون حجم العيّنة كافياً ، يمكننا الحصول انطلاقاً منها على تقريب مناسب للنسبة أو للمتوسّط في مجمل المجتمع الإحصائي : يشكّل قانون الأعداد الكبيرة أساس طريقة الأبحاث الإحصائية .

لقانون الأعداد الكبيرة شروط تطبيق عامّة جدّاً لأنّه لا يستدعي إدخال قانون احتمال المتغيّرة موضع الدراسة . وهو يستند بالمقابل إلى سلسلة من العلاوات المهمّة (majorations) (لا مساواة بيانيميه ـ تشيبيتشيف) ويؤدّي إلى مقادير عيّنة أكبر بكثير ، في الحقيقة من أن تكون ضرورية للحصول على الدقّة المطلوبة . طبعاً من المفضّل أن نحسب مباشرة حجم العيّنة انطلاقاً من قانون الاحتمال عندما يكون هذا الأمر ممكناً .

مثلاً . في المثل السابق (ص 233) ، عيّنة من 000 240 وحدة إحصائية هي ترف غير مفيد للحصول ، باحتمال 99% ، على تقدير لـ p بفارق 1/100 :

$$P\{\,|f_n - p| \leqslant 0{,}01\,\} \geqslant 0{,}99 .$$

في الواقع ، في هذه الحالة ، نحن نعرف توزيع احتمال التردّد f_n : إنّه قانون ذو حدّين بمتغيّرين وسيطين n و p=0,04 . وبما أنّ حجم العيّنة n هو حتماً كبير بشكل كاف ، يمكننا تقريب هذا القانون من قانون طبيعي (معتدل) بمتغيّرين وسيطين $m = p$ و $\sigma = \sqrt{pq/n}$. بالتالي :

- يمكننا تحديد قيمة المتغيّرة الطبيعية المركزة المختصرة t بشكل يتواجد معه 99 فرصة على 100 كي تكون f_n ضمن الفسحة $p \pm t\sqrt{pq/n}$:

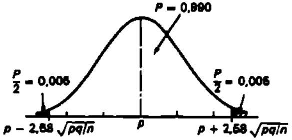
$$P\left\{p - t\sqrt{pq/n} \leqslant f_n < p + t\sqrt{pq/n}\right\} \geqslant 0{,}99 .$$

تعطينا مراجعة الجدول P(t) : t = 2,58

- بعد تثبيت قيمة t وكي نحصل على :

$$t\sqrt{pq/n} \leqslant 0{,}01$$

$$2{,}58\sqrt{0{,}4 \times 0{,}6/n} \leqslant 0{,}01 ,$$



يكفي أن نختار :

$$n \geqslant 15\,975 \neq 16\,000 .$$

إذن من غير المفيد أن نعمد إلى 000 240 مشاهدة لأن 16000 (أي 15 مرّة أقلّ) تكفي للحصول على الدقّة المطلوبة .

الفصل السادس

تأويل الأبحاث الإحصائية العشوائية :
مسائل التقدير والمقارنة

لا يمكن أن يتكوّن لدى المشرف على مشروع معيّن يقين مطلق بالنسبة للدقّة حول المعلومات المحصودة من البحث الإحصائي : إذ يبقى قسم من المصادفة ملازماً لهذه الطريقة . إذاً المسألة التي تطرح نفسها ، إنطلاقاً من المشاهدات على العيّنة ، هي مسألة تقدير هذا المقياس أو ذاك من مقاييس المجتمع الإحصائي وتقييم دقّة هذا التقدير وذلك مع أقصى ما يمكن من الفعالية .

هكذا ، فيما يخصّ دراسات العرض والطلب ، يرغب المسؤول عن توزيع مادّة استهلاكية مهمّة في الحصول ، انطلاقاً من حملة البحث الإحصائي ، على معدّل نفقات مختلف فئات الشعب على هذا النوع من المشتريات ؛ وينوي صانع للسيارات أن يقدّر نسبة الأسر التي تملك سيارة وتوزيع هذه السيارات حسب الماركة ، العمر ، . . . في ما يخصّ فحص النوعية ، قد نرغب عند استلامنا كمّية من القطع الميكانيكية بتقييم نسبة النفاية التي تتركها الكمية (القطع المعيبة) ؛ وتسمح لنا عملية جرد كمّية من المصنوعات بواسطة البحث الإحصائي بتقدير النسبة المئوية للأخطاء المرتكبة عند إجراء العملية ، الخ . .

ولبعض المسائل التي نصادفها عملياً طبيعة أخرى : فالأمر يتعلّق بالمقارنات أكثر منه بالتقديرات . مثلاً ، عند استلام كمّية من القطع المصنوعة بالجملة ، قد نهتمّ بمقارنة نسبة النفاية الملحوظة مع الحدّ في العقد ، كي نرفض الكمّية عند تجاوز هذا الحد ، أكثر من اهتمامنا بالتقدير غير الدقيق حتماً لنسبة نفايات الكمّية . كذلك ، بعد حملة تنمية مبيعات اعتمدت طريقتين مختلفتين ، يرغب مدير المشروع التجاري بتحديد الطريقة

الأكثر فعالية ، دون أن يطمح لإعطاء رقم دقيق للمردودين الحاصلين .

القسم I

مسائل التقدير

1 . المقدّرات : A . مفهوم المقدّر ؛ B . مقدّرات المقاييس الرئيسية للمجتمع الإحصائي . ـ2 . فسحة الثقة للتقدير : A . تقدير المتوسّط ؛ B . تقدير النسبة ؛ C . تحديد حجم العيّنة .

حول موضوع تقدير أحد مقاييس المجتمع المرجع انطلاقاً من عيّنة عشوائية ، ينطرح نوعان من المسائل .

ينبغي أوّلاً البحث عن الكمية ، المحسوبة على العيّنة ، القادرة على أن تعطينا بشكل صحيح وفعّال تقديراً للمقياس المقصود : إنّه اختيار المقدّر .

يجب بعدئذٍ تحديد دقّة التقدير بإحاطتنا الرقم الذي نحصل عليه مساحة من القيم وبإعطائنا حجم المخاطرة لوجود القيمة الحقيقية خارج هذه المساحة : إنّه تحديد فسحة ثقة التقدير .

1 . المقدّرات

لنفترض أنّ جهازاً للدراسات الاقتصادية استنتج بعد أخذه عيّنة من n = 10 000 أسرة ، أنّ القيمة المتوسّطة للنفقات المخصّصة للسكن تبلغ :

$$\bar{x} = 200 \text{ F}$$

كيف نقدّر انطلاقاً من هذه النتيجة متوسّط نفقة السكن m في المجتمع الإحصائي ككلّ ؟ إنّ متوسّط العيّنة هو ، قبل تحديدها ، متغيّرة عشوائية $\bar{X}$ أملها الرياضي m و(في حالة المسحوبات المستقلّة) انحرافها النموذجي (أو المعياري) $\sigma/\sqrt{n}$.

بفضل قانون الأعداد الكبيرة ، تميل $\bar{X}$ بالاحتمال نحو القيمة الحقيقية m لمتوسّط نفقة السكن في المجتمع الإحصائي عندما يتزايد مقدار العيّنة n بشكل غير متناه .

يبدو أنّه من الطبيعي إذاً أن نعتمد متوسّط العيّنة $\bar{X}$ كمقدّر لِـ m . القيمة الملحوظة ، $\bar{x} = 200$ F ، هي تقدير m الحاصل انطلاقاً من هذه العيّنة بالذات .

A . مفهوم المقدّر

تعريف

لنفترض أنّنا نريد تقدير المقياس Θ للمجتمع المرجع ، مثلاً : متوسّـط المتغيّـرة X أو تباينها .

لنفتـرض أنّ x_1 ، x_2 ، . . . ، x_n هي القيم التي تأخـذها X بـالنسـبة لـوحدات العيّـنة وأنّ $\theta(x_1, x_2, ... x_n)$ هي دالّـة حسب هذه القيم .

θ ، كونها دالّـة حسب المتغيّـرات العشوائية x_1 ، x_2 ، . . . ، x_n ، هي نفسها متغيّـرة عشوائية تأخذ قيمة معيّـنة لكلّ عيّـنة .

نقول أنّ $\theta(x_1, x_2, ..., x_n)$ هي مقدِّر لـ Θ إذا كان :

$$E\{\theta\} \to \Theta$$

$$V\{\theta\} \to 0$$

عندما تتزايد n بصورة غير متناهية .

بعبارة أخرى ، تعتبر θ مقدِّراً لـ Θ إذا كان يكفي اختيار مقدار العيّـنة n كبيراً بدرجة كافية بجعل قانون توزيع θ منحصراً حول Θ قدر ما نريد . ونتمسّـك بهذه الخاصّـة بقولنا أنّ θ هي مقدِّر متقارب لـ Θ .

نأخذ قيمة θ العددية الملحوظة على العيّـنة الوحيدة كتقدير لـ Θ . حول هذا التقدير الموضعي نحدّد فسحة ثقة تعطينا درجة خطأ المعاينة الذي قد يُرتكب .

نوعية المقدِّر

نميّـز المقدِّر الجيّـد بغياب تحيّـزه وضعف تشتّـه .

أ ـ المقدِّر غير المتحيّـز

نقول أنّ المقدّر θ هو غير متحيّـز (أو غير منحرف) إذا كان : $E\{\theta\} = \Theta$.

يكون المقدِّر عندئذٍ مركزاً عنـد قيمة Θ الحقيقيـة ، مهما كـان مقدار العيّـنـة .

التحيّـز $B(\theta)$ يـساوي :

$$B(\theta) = E(\theta) - \Theta .$$

والتحيّـز هـو خطأ منهجي ، ورغم مسـاوىء هذا الخـطأ قد يكـون لمن الأفضل

استعمال مقدّر متحيّز بشكل طفيف إذا كان تشتّته أضعف من تشتّت مقدّر غير متحيّز ، ولكن تجهل معرفة حدّ أعلى للتحيّز . ويحكم تقارب (ميل) المقدّر ، يمكن تصغير التحيّز قدر ما نريد بتكبيرنا حجم العيّنة :

$B(\theta) = [E\{\theta\} - \Theta] \rightarrow 0$ عندما تتزايد n بصورة غير متناهية .

ب ـ المقدّر ذو التشتّت الضعيف

يكون المقدّر θ أفضل قدر ما يتضمّن خطأ عشوائياً أضعف . وتُقاس قابلية تغيّر θ بواسطة تباينها :

$$V(\theta) = E\{(\theta - E\{\theta\})^2\}.$$

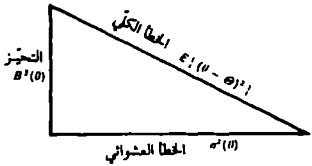
بين مقدرين غير متحيّزين ، الأكثر فعالية هو ، تعريفاً ، الذي يملك التباين الأصغر.

يتألّف الخطأ المنهجي والخطأ العشوائي كما ضلعا الزاوية القائمة في مثلّث قائم الزاوية كي يعطيا الخطأ الكلّي :

$$\begin{aligned} E\{(\theta - \Theta)^2\} &= E\{(\theta - E\{\theta\} + E\{\theta\} - \Theta)^2\}, \\ &= E\{(\theta - E\{\theta\})^2\} + (E\{\theta\} - \Theta)^2, \\ &= \sigma^2(\theta) + B^2(\theta) \end{aligned}$$

في الواقع :

$$E\{(\theta - E\{\theta\})(E\{\theta\} - \Theta)\} = E[\theta - E\{\theta\}](E\{\theta\} - \Theta) = 0$$



عندما نسحب عيّنة واحدة ، وهذا هو الحال الأكثر مصادفة ، لا داعي للتمييز بين التحيّز والخطأ العشوائي : الخطأ الكلّي هو الذي يؤخذ بعين الاعتبار . قد يكون في صالحنا إذن استعمال مقدّر متحيّز بعض الشيء كي نتقدّم على الخطأ العشوائي .

B . مقدّرات المقاييس الرئيسية للمجتمع الإحصائي

لنأخذ مجتمعاً إحصائياً مؤلّفاً من N وحدة U_s نعاينها بواسطة رقمها s :

$$s = 1, 2, \ldots, N$$

ونسحب من هذا المجتمع عيّنة مقدارها n ، حيث نتعرّف إلى وحدات هذه العيّنة U_i بواسطة رتبتها i خلال السحب :

$$i = 1, 2, \ldots, n$$

الرموز

لنأخذ المتغيّرة X ..

سوف نرمز في المجتمع الإحصائي :

إلى قيمة المتغيّرة X للوحدة الإحصائية U_s بواسطة X_s ،

إلى متوسّط X بواسطة m :

$$m = \frac{1}{N}\sum_{s=1}^{N} X_s .$$

إلى تباين X بواسطة σ^2 :

$$\sigma^2 = \frac{1}{N}\sum_{s=1}^{N} (X_s - m)^2 .$$

وفي العيّنة ، سنشير إلى الكمّيات المشابهة بواسطة :

x_i للدلالة على قيمة المتغيّرة X لوحدة العيّنة U_i ،

$\bar{x}$ للدلالة على متوسّط X :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i .$$

s^2 للدلالة على تباين X :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

لقد بدا لنا مسبقاً أنّه من الطبيعي تقدير متوسّط المجتمع الإحصائي m بواسطة متوسّط العيّنة $\bar{x}$ المسحوبة من هذا المجتمع . لنعد إلى هذه النقطة بحسابنا بشكل أدق أمل $\bar{x}$ الرياضي وتباينها تبعاً لطريقة سحب العيّنة .

أ ـ الأمل الرياضي والتباين لمتوسّط العيّنة

1 . العيّنة المستقلّة (المسحوبة مع ردّ)

إنّ x_i ، وهي قيمة المتغيّرة X بالنسبة لوحدة العيّنة المختارة عند السحب رقم i ، هي متغيّرة عشوائية يمكنها أخذ واحدة من القيم التالية :

$$X_1, X_2, \ldots, X_\alpha, \ldots, X_N$$

باحتمال يساوي 1/N .

إذاً ، أملها الرياضي يساوي متوسّط المجتمع الإحصائي m :

$$E\{x_i\} = \frac{1}{N}\sum_{\alpha=1}^{N} X_\alpha = m$$

وتباينها يساوي تباين المجتمع الإحصائي

$$V\{x_i\} = \frac{1}{N}\sum_{\alpha=1}^{N}(X_\alpha - m)^2 = \sigma^2$$

أمل متوسّط العيّنة الرياضي

بناء على تعريف الأمل الرياضي :

$$E\{\bar{x}\} = E\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n}\sum_{i=1}^{n} E\{x_i\}.$$

وذلك بفضل خصائص الأمل الرياضي (أنظر الفصل I ، ص 56) . بالتالي :

$$E\{\bar{x}\} = m.$$

لأنّه ، كما أثبتنا لتوّنا :

$$E\{x_i\} = m.$$

تباين متوسّط العيّنة

بناء على تعريف التباين :

$$V\{\bar{x}\} = V\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} V\{x_i\}.$$

وذلك بفضل استقلالية السحوبات وخصائص التباين (أنظر الفصل I ، ص 61) . بالتالي :

$$V(\bar{X}) = \frac{\sigma^2}{n}.$$

لأنّـه ، كما أثبتنا لتوّنا :

$$V(X_i) = \sigma^2.$$

2 . العيّـنة المستنفِدة (المسحوبة دون ردّ)

كي نحسب أمل متوسّـط العيّـنة المستنفِدة الرياضي وتبـاينه ، سـوف نعمد إلى حيلة في الحساب تعود إلى كورنفيلد (Cornfield) .

إلى كـلّ وحدة U_i من المجتمـع الإحصـائي ، نـنسب متغيِّرة بـرنولي ε_i التي نعطيها القيمة 1 إذا كانت U_i تنتمي إلى العيّـنة E ، وصفر في الحالة المعاكسة . إنّ قانون احتمال هذه المتغيِّـرة هو التالي :

الحدث	المتغيِّـرة العشوائية ε_i	الاحتمال $P\{\varepsilon_i\}$
$U_i \in E$	1	n/N
$U_i \notin E$	0	$1 - n/N$

ففي الواقع ، الاحتمال p_i لأن تنتمي الوحدة U_i إلى العيّـنة :

$$p_i = P\{\varepsilon_i = 1\}.$$

هو نفسـه مهما كانت الوحدة المأخوذة بعين الاعتبار . من جهـة أخرى وبنـاء على تعريف ε_i :

$$\sum_{i=1}^{N} \varepsilon_i = n.$$

بالتالي :

$$E(n) = n = \sum_{i=1}^{N} E(\varepsilon_i) = \sum_{i=1}^{N} p_i = N.p_i \quad [1]$$

إذاً :

$$p_i = \frac{n}{N}.$$

(1) بما أنّ n هو عدد ثابت ، E (n) = n . من جهة أخرى ، بفضل خصائص الأمل الرياضي :

$$E\left\{\sum_{i=1}^{N} \varepsilon_i\right\} = \sum_{i=1}^{N} E(\varepsilon_i)$$

وبناء على التعريف :

$$E(\varepsilon_i) = 1.p_i + 0(1 - p_i) = p_i.$$

وإذا استعملنا هذه المتغيّرة المؤشّرة ε_s، باستطاعة متوسّط العيّنة :

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

أن يُكتب :

$$\bar{x} = \frac{1}{n}\sum_{s=1}^{N} X_s.\varepsilon_s .$$

في هذه العبارة ، القيم X_s هي أعداد ثابتة ؛ وحدها القيم ε_s هي متغيّرات عشوائية وضعنا لتوّنا قانون احتمالها .

أمل متوسّط العيّنة الرياضي

بناء على التعريف :

$$E\{\bar{x}\} = E\left\{\frac{1}{n}\sum_{s=1}^{N} X_s.\varepsilon_s\right\} = \frac{1}{n}\sum_{s=1}^{N} X_s E\{\varepsilon_s\}.$$

ولكن ، بحكم تعريف الأمل الرياضي :

$$E\{\varepsilon_s\} = 1.\frac{n}{N} + 0\left(1-\frac{n}{N}\right) = \frac{n}{N}.$$

إذاً :

$$E\{\bar{x}\} = \frac{1}{n}.\frac{n}{N}\sum_{s=1}^{N} X_s = \frac{1}{N}\sum_{s=1}^{N} X_s = m.$$

تباين متوسّط العيّنة

بناء على التعريف :

$$V\{\bar{x}\} = E(\bar{x}-m)^2 .$$

وإذا استعملنا المتغيّرة المؤشّرة ، بإمكاننا أن نكتب :

$$\bar{x} - m = \frac{1}{n}\sum_{s=1}^{N}(X_s - m).\varepsilon_s$$

$$(\bar{x}-m)^2 = \frac{1}{n^2}\sum_{s=1}^{N}(X_s-m)^2.\varepsilon_s^2 + \frac{1}{n^2}\sum_{\substack{s=1\\ s\neq s'}}^{N}\sum_{s'=1}^{N}(X_s-m)(X_{s'}-m).\varepsilon_s\varepsilon_{s'}$$

إذاً ، بفضل خصائص الأمل الرياضي :

$$V\{\bar{x}\} = E(\bar{x}-m)^2 = \frac{1}{n^2}\sum_{s=1}^{N}(X_s-m)^2.E\{\varepsilon_s^2\}$$

$$+\frac{1}{n^2}\sum_{s=1}^{N}\sum_{s'=1}^{N}(X_s-m)(X_{s'}-m).E\{\varepsilon_s\varepsilon_{s'}\},$$

حيث $(X_s-m)^2$ و $(X_{s'}-m)^2$ هما كميتان ثابتتان .

حساب $E(\varepsilon_s^2)$

بناء على تعريف الأمل الرياضي :

$$E(\varepsilon_s^2) = 1 . \frac{n}{N} + 0\left(1-\frac{n}{N}\right) = \frac{n}{N}.$$

حساب $E(\varepsilon_s \varepsilon_{s'})$

إنّ حاصل الضرب $\varepsilon_s \varepsilon_{s'}$ يساوي 1 عندما تنتمي الوحدتان U_s و $U_{s'}$ معاً إلى العيّنة. احتمال هذا الحدث $p_{ss'}$ يساوي :

$$\frac{n}{N} \cdot \frac{n-1}{N-1}.$$

ففي الواقع ، بتطبيقنا لقاعدة الاحتمالات المركّبة :

$$p_{ss'} = p_s \cdot p_{s'/s} ,$$

حيث يعبّر $p_{s'/s}$ عن الاحتمال الشرطي لأن تنتمي $U_{s'}$ إلى العيّنة مع العلم أنّ U_s انتمت إليها . وقد رأينا أعلاه أنّ :

$$p_s = \frac{n}{N}.$$

وإذا اتّبعنا نمط تفكير مشابه بعد أن نطرح U_s من المجتمع الإحصائي ومن العيّنة ، نحصل على :

$$p_{s'/s} = \frac{n-1}{N-1}.$$

ويساوي حاصل الضرب $\varepsilon_s \varepsilon_{s'}$ صفراً في كلّ الحالات الأخرى . بالتالي :

$$E(\varepsilon_s \varepsilon_{s'}) = 1 . \frac{n}{N} \cdot \frac{n-1}{N-1} + 0\left(1-\frac{n}{N} \cdot \frac{n-1}{N-1}\right) = \frac{n}{N} \cdot \frac{n-1}{N-1}.$$

لنضع هذه النتائج في عبارة $V(\bar{x})$:

$$V(\bar{x}) = \frac{1}{n} \cdot \frac{1}{N} \sum_{s=1}^{N} (X_s - m)^2 + \frac{1}{n} \cdot \frac{n-1}{N-1} \cdot \frac{1}{N} \sum_{s=1}^{N} \sum_{\substack{s'=1 \\ s' \neq s}}^{N} (X_s - m)(X_{s'} - m).$$

وهذا يمكننا كتابته ، إذا وضعنا $\frac{1}{n} \cdot \frac{n-1}{N-1} \cdot \frac{1}{N}$ كعامل مشترك :

$$V\{\bar{x}\} = \frac{1}{n}\cdot\frac{n-1}{N-1}\cdot\frac{1}{N}\left[\sum_{i=1}^{N}(X_i - m)^2 + \sum_{i=1}^{N}\sum_{\substack{j=1\\i\neq j}}^{N}(X_i - m)(X_j - m)\right]$$

$$+ \frac{1}{n}\left(1 - \frac{n-1}{N-1}\right)\frac{1}{N}\sum_{i=1}^{N}(X_i - m)^2 .$$

إلاّ أنّ :

$$\sum_{i=1}^{N}(X_i - m)^2 + \sum_{i=1}^{N}\sum_{\substack{j=1\\i\neq j}}^{N}(X_i - m)(X_j - m) = \left[\sum_{i=1}^{N}(X_i - m)\right]^2 = 0 .$$

لأنّ :

$\sum_i (X_i - m) = 0.$ بناء على تعريف m .

من ناحية أخرى :

$$\frac{1}{N}\sum_{i=1}^{N}(X_i - m)^2 = \sigma^2 .$$

إذاً :

$$V\{\bar{x}\} = \frac{N-n}{N-1}\cdot\frac{\sigma^2}{n} .$$

ونقارب نتيجة التباين هذه مع تباين النسبة f للوحدات التي تمثّل الخاصّة A ، الملحوظة على عيّنة مستنفدة (المتغيّرة فوق الهندسية ، انظر الفصل II ، القسم II ، ص 86) :

$$V\{f\} = \frac{N-n}{N-1}\cdot\frac{pq}{n} .$$

في الواقع ، كما سنرى لاحقاً (مقدّر النسبة ، ص 250) ، يمكننا دوماً النظر إلى النسبة كمتوسّط متغيّرة برنولي يساوي تباينها pq .

بالاختصار :

- يساوي أمل متوسّط العيّنة $\bar{x}$ الرياضي متوسّط المجتمع الإحصائي m الذي سُحبت منه هذه العيّنة ، مهما كانت طريقة السحب :

$$E\{\bar{x}\} = m ;$$

- يساوي تباين $\bar{x}$ ، في حالة العيّنة المستقلّة :

$$V\{\bar{x}\} = \frac{\sigma^2}{n}$$

وفي حالة العيّنة المستنفدة :

$$V(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}.$$

العامل (N-n)/(N-1) الذي يصغّر ، في حالة السحب المستنفِد ، تباين المقدِّر تبعاً لمقدار العيّنة ، يُدعى مُعامِل الاستنفاد .

ب ـ المقدّرات الرئيسية

1 . مقدّر متوسّط المجتمع الإحصائي

نستنتج ممّا سبق أنّ المتوسّط $\bar{x}$ ، الملحوظ على العيّنة هو ، مهما كان نوع طريقة السحب ، مقدّر غير متحيّز لمتوسّط المجتمع الإحصائي :

$$E(\bar{x}) = m.$$

تباين هذا المقدّر هو :

في حالة السحوبات المستقلّة : $V(\bar{x}) = \frac{\sigma^2}{n}$

في حالة السحوبات المستنفِدة : $V(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$

كون مُعامِل الاستنفاد (N - 1) / (N - n) دائماً أصغر من 1 ، فإنّه عندما يكون الحجم نفسه ، يعتبر متوسّط عيّنة مستنفدة مقدّراً أكثر فعالية لمتوسّط المجتمع الإحصائي من متوسّط عيّنة مستقلّة .

غالباً ما يكون مقدار المجتمع الإحصائي عدداً مرتفعاً . بالتالي قليلاً ما يختلف المعامل (N-n)/(N-1) عن 1-n/N الذي يمثّل المتمّم إلى واحد لنسبة البحث الإحصائي t = n/N . لدينا :

$$V(\bar{x}) \neq \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right).$$

من جهة أخرى ، عندما يكون مقدار العيّنة n ضعيفاً بالنسبة لمقدار المجتمع الإحصائي N ، يمكننا إهمال العبارة (N-n)/(N-1) التي تقترب قيمتها من 1 . بالتالي ، عندما تكون نسبة البحث الإحصائي ضعيفة ، تكون طريقتا السحب تقريباً متعادلتين ولا تتوقّف دقّة التقديرات ، كتقريب أوّل ، إلاّ على مقدار العيّنة ، وليس على نسبة البحث . تُعتبر هذه النتيجة مهمّة لأنّها تُظهر أنّ كلفة البحث الإحصائي تكون أكبر كلّما كان المجتمع الإحصائي موضوع الدراسة أصغر .

2 . مقدّر تباين المجتمع الإحصائي

كي نقدّر تباين المجتمع الإحصائي :

$$\sigma^2 = \frac{1}{N}\sum_{s=1}^{N} (X_s - m)^2$$

يخطر لنا لأوّل وهلة أن نستعمل ، كما بالنسبة للمتوسّط ، الكمّية المطابقة ، أي التباين المقاس على العيّنة :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

إلاّ أنّ هذا المقدّر متحيّز .

لنحسب في الواقع :

$$E\{s^2\} = E\left\{\frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2\right\} .$$

وإذا أدخلنا متوسّط المجتمع الإحصائي m يمكننا أن نكتب ، انطلاقاً من النتيجة الموضوعة في « الإحصاء الوصفي » ، القسم I ، الفقرة 3 C :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2 - (\bar{x} - m)^2 .$$

بالتالي :

$$E\{s^2\} = \frac{1}{n}\sum_{i=1}^{n} E(x_i - m)^2 - E(\bar{x} - m)^2$$

$$= V\{x\} - V\{\bar{x}\} . \qquad (1)$$

ـ عيّنة مستقلّة (مسحوبة مع ردّ)

$$V\{x\} = \sigma^2 \qquad V\{\bar{x}\} = \frac{\sigma^2}{n}$$

إذاً ، إذا انتقلنا إلى (1) :

$$E\{s^2\} = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2 .$$

بالتالي ، المقدّر غير المتحيّز لتباين المجتمع ليس s^2 ، بل :

$$s'^2 = \frac{n}{n-1}s^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

وقد أتى الالتواء في هذا الحساب نتيجة قياس الانحرافات بالنسبة إلى متوسّط العيّنة وليس بالنسبة إلى متوسّط المجتمع الإحصائي .

ومقدَّر تباين $\bar{x}$ هو ، بعد أن نستبدل σ^2 بتقديرها من خلال العيّنة :

$$V^*(\bar{x}) = \frac{s'^2}{n}$$

- عيّنة مستنفِدة (مسحوبة دون ردّ)

$$V(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}$$

إذاً ، إذا انتقلنا إلى (1) :

$$E(s^2) = \sigma^2 - \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n} = \frac{N}{n} \cdot \frac{n-1}{N-1} \sigma^2$$

بالتالي ، المقدَّر غير المتحيِّز لتباين المجتمع الإحصائي ليس s^2 ، بل :

$$\frac{n}{N} \cdot \frac{N-1}{n-1} s^2 = \frac{N-1}{N} \cdot \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{N-1}{N} s'^2$$

ومقدَّر تباين $\bar{x}$ هو ، بعد أن نستبدل σ^2 بمقدَّرها من خلال العيّنة :

$$V^*(\bar{x}) = \frac{N-n}{N-1} \cdot \frac{N-1}{N} \cdot \frac{s'^2}{n} = \frac{N-n}{N} \cdot \frac{s'^2}{n}$$

بالاختصار ، يُقدَّر تباين متوسّط العيّنة بواسطة :

في حالة السحوبات المستقلّة : $$V^*(\bar{x}) = \frac{s'^2}{n}$$

في حالة السحوبات المستنفِدة : $$V^*(\bar{x}) = \frac{N-n}{N} \frac{s'^2}{n}$$

في هاتين العبارتين ، s'^2 ترمز إلى المقدَّر غير المتحيِّز لتباين المجتمع الإحصائي إنطلاقاً من عيّنة مستقلّة :

$$s'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

مع ذلك ، عندما يكون مقدار العيّنة n كبيراً ، لا يكون s'^2 مختلفاً كثيراً عن التباين s^2 المقاس على العيّنة :

3 . مقدّر النسبة

لنأخذ مجتمعاً إحصائياً يتضمّن فئتين من الوحدات :

- الوحدات A بنسبة p ،
- الوحدات B بنسبة $q = 1 - p$

يمكننا اعتبار النسبة p كمتوسّط m لمتغيّرة برنولي تأخذ القيمة 1 بالنسبة للوحدات A والقيمة صفر بالنسبة للوحدات B :

$$m = p.1 + q.0 = p$$

يمكننا إذاً إرجاع تقدير النسبة p إنطلاقاً من عيّنة ما إلى تقدير متوسّط من نوع خاص (أنظر الفصل III ، ص 115) . ونأخذ كمقدّر لهذه الكمّية التردّد f للوحدات A في العيّنة ، أي متوسّط المتغيّرة X الملحوظ على العيّنة .

يساوي تباين X :

$$\sigma^2 = p(1 - p)^2 + q(0 - p)^2 = pq^2 + qp^2 = pq(p + q) = pq .$$

تباين المقدّر هو إذاً :

- في حالة السحوبات المستقلّة :

$$V\{f\} = pq/n .$$

وهنا نتعرّف إلى عبارة تباين التردّد ذي الحدّين (أنظر الفصل II ، القسم I ، ص 77) ؛

- في حالة السحوبات المستنفِدة :

$$V\{f\} = \frac{N - n}{N - 1}.\frac{pq}{n};$$

f هو ، في الواقع ، في هذه الحالة تردّد فوق هندسي حيث نتعرّف إلى عبارة تباينه (أنظر الفصل II ، القسم II ، ص 86) .

ويساوي تباين X مقاساً على العيّنة :

$$s^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$= \frac{1}{n}[nf(1 - f)^2 + n(1 - f)(0 - f)^2] = f(1 - f)$$

لأنـه ، في العيّـنة وبناء على تعريف التردّد (f= X/n) ، تأخذ المتغيّـرة X ، nf مرّة القيمة 1 و n(1−f) مرّة القيمة 0 .

يمكننا إذن تقدير pq ، وهي تباين X في المجتمع الإحصائي ، بواسطة :

$$s'^2 = \frac{n}{n-1} s^2 = \frac{n}{n-1} f(1-f).$$

بالاختصار ، نختار f ، وهو التردّد الملحوظ على العيّـنة ، كمقـدِّر لِـ p . ويُقدّر تباين هذا المقدِّر بواسطة :

في حالة السحوبات المستقلّة : $V^*\{f\} = \frac{f(1-f)}{n-1}$

في حالة السحوبات المستنفِدة : $V^*(f) = \frac{N-n}{N} \frac{f(1-f)}{n-1}$

4 . مقدِّر المجموع

بحكم تعريف المتوسّـط m :

$$S = \sum_{i=1}^{N} X_i = Nm.$$

نختار كمقدِّر للمجموع s الكمّية $N\bar{x}$ وتباينها :

$$V\{N\bar{x}\} = N^2 V\{\bar{x}\},$$

الذي نقدّره بواسطة :

$$V^*\{N\bar{x}\} = N^2 V^*\{\bar{x}\}.$$

5 . مقدِّر المقدار

المقـدار N_A للوحدات A الـموجودة في المجتمـع الإحصائي يسـاوي Np . نختار كمقدِّر له N_f التي نقدِّر تباينها : $V\{Nf\} = N^2 V\{f\}$ بواسطة :

$$V^*\{Nf\} = N^2 V^*\{f\}.$$

ونكتب مُعامِل التغيّـر CV ، الذي يقيس دقّـة التقدير :

$$(CV)^2 = \frac{V\{Nf\}}{(Np)^2} = N^2 \frac{pq}{n} \frac{1}{(Np)^2} = \frac{q}{np}$$

في حال عيّـنة مستقلّة أو ، في حالة عيّـنة مستنفِدة ، بإهمالنا المعامل التصحيحي (مقدار العيّـنة n ضعيف بالنسبة لمقدار المجتمع الإحصائي) .

إذا كان المجتمع الثانوي الذي نسعى إلى تقديره قليلاً نسبياً ، لا تختلف q كثيراً عن 1 : ونحصل على عبارة قريبة من مُعامِل تغيّـر بسيط :

$$(CV)^2 \approx \frac{1}{np}.$$

عندئذٍ لا تتوقّـف دقّـة التقدير إلّا بـ np الذي يمثّـل الأمل الرياضي لعدد وحدات العيّـنة التي تنتمي إلى الفئة التي نسعى إلى تقدير مقدارها .

2 . فسحة ثقة التقدير

لقد رأينا كيف يمكننا ، انطلاقاً من العيّـنة ، تقدير المقاييس الرئيسية للمجتمع الإحصائي . يبقى أن نحدّد دقّـة هذه التقديرات .

لنفترض أن Θ هو مقياس المجتمع الإحصائي الذي يجب تقديره ، و θ هو مقدِّره انطلاقاً من العيّـنة .

لنتّفق على قيمة احتمال معيّـن α ، مثلاً α =5% : نقبل تحمّـل مخاطرة باحتمال α =5% لأن نرتكب خطأً بالنسبة لدقّـة التقدير .

بمعرفتنا قانون احتمال المقدِّر θ ، يمكننا تحديد الفسحة $(\Theta - h_1, \Theta + h_2)$ حول قيمة Θ الحقيقية بشكل يكون فيه للكمّية θ الملحوظة على العيّـنة الاحتمال 1 − α للانتماء إلى هذه الفسحة :

$$P\{\Theta - h_1 \leqslant \theta \leqslant \Theta + h_2\} = 1 - \alpha.$$

عدم المساواة المزدوج :

$$\Theta - h_1 \leqslant \theta \leqslant \Theta + h_2$$

يعادل :

$$\theta - h_2 \leqslant \Theta \leqslant \theta + h_1.$$

ننسب إذن إلى الفسحة $(\theta - h_2, \theta + h_1)$ الاحتمال 1−α لأن نغطّي قيمة

θ الحقيقية المجهولة :

$$P\{\hat{\theta} - h_2 \leqslant \theta \leqslant \hat{\theta} + h_1\} = 1 - \alpha .$$

وتسمّى هذه الفسحة بفسحة ثقة تقدير θ بدرجة الاحتمال $1-\alpha$: إذا كانت $\alpha = 5\%$ هناك 95 فرصة على 100 أن توجد قيمة θ الحقيقية في الفسحة المحدّدة بهذه الطريقة حول القيمة الملحوظة $\hat{\theta}$.

ويكون المقدّر أكثر فعالية كلّما أدّى ، بالنسبة لدرجة احتمال $1-\alpha$ معيّنة ، إلى فسحة ثقة أصغر .

A . تقدير المتوسّط

1 . المتوسّط $\bar{x}$ لعيّنة مأخوذة من مجتمع إحصائي موزّع طبيعياً يتوزّع هو نفسه طبيعياً .

بشكل عام أكثر ، يمكننا تشبيه توزيع المتوسّط $\bar{x}$ لعيّنة مأخوذة من أيّ مجتمع إحصائي متوسّطه m وانحرافه النموذجي σ بقانون طبيعي متوسّطه m وانحرافه النموذجي $\sigma_{\bar{x}}$ ، وذلك منذ أن يتجاوز مقدار العيّنة الثلاثين وحدة (أنظر الفصل III، ص 113) :

$$\bar{x} = \mathcal{N}(m, \sigma_{\bar{x}}) .$$

في حالة عيّنة مسحوبة مع ردّ :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} .$$

وفي حالة عيّنة مستنفِدة :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \approx \frac{\sigma}{\sqrt{n}} \sqrt{1 - \frac{n}{N}} ,$$

حيث n/N يمثّل نسبة البحث الإحصائي .

2 . بشكل عام ، يكون انحراف المجتمع الإحصائي النموذجي σ مجهولاً ، كشأن m . عندئذٍ نستعمل تقديره s' المستنتج من المشاهدات :

$$s'^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

عندما يكون مقدار العيّنة مرتفعاً ، لا يختلف هذا التقدير كثيراً عن قيمة الانحراف النموذجي المسحوب على العيّنة :

$$s'^2 \neq s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

ومقدّر $\sigma_{\bar{x}}$ هو (أنظر الفقرة 1.B ، ص 249) :

في حالة السحوبات المستقلّة : $\frac{s'}{\sqrt{n}}$

في حالة السحوبات المستنفِدة : $\frac{s'}{\sqrt{n}}\sqrt{1-\frac{n}{N}}$

- إذا كان مقدار العيّنة كبيراً (أكبر من 30) ، s'^2 هو تقدير لِـ σ^2 دقيق كفاية لكي تكون المتغيّرة المركزة المختصرة التالية ، حيث استبدلنا σ لحسابها بواسطة s' :

(سحوبات مستقلّة) $T = \frac{\bar{x} - m}{s'/\sqrt{n}}$

أو (سحوبات مستنفِدة) $. T = \frac{\bar{x} - m}{\frac{s'}{\sqrt{n}}\sqrt{1-\frac{n}{N}}}$

موزّعة حسب القانون الطبيعي (المعتدل) .

- إذا كان مقدار العيّنة n صغيراً (أقلّ من 30) ، فإنّه نتيجة تقلّبات مخرج T العشوائية ، لا يمكن تشبيه هذه المتغيّرة بمتغيّرة طبيعية مركزة مختصرة . فحسب الافتراض المقيّد لعيّنة غير مستنفِدة مأخوذة من مجتمع إحصائي طبيعي ، تتبع هذه المتغيّرة قانون ستودنت ـ فيشر (Student-Fisher) ذا n−1 درجة حرّية ، وقد تمّ حساب جداول هذا القانون (الملحق : الجدول 6) .

بالاختصار ، في الحالة التي تُصادف غالباً ، حالة العيّنة الكبيرة (بمقدار أكبر من 30 وحدة إحصائية) لا نلتقي أثناء تحديدنا لفسحة ثقة تقدير المتوسّط بصعوبات تذكر : مهما كان التوزيع الأصل فإنّ متوسّط العيّنة يتبع قانوناً طبيعياً يمكننا تقدير انحرافه النموذجي إنطلاقاً من العيّنة نفسها .

مثل 1 . سحبنا عيّنة مستنفِدة تتألّف من 10 000 أسرة في منطقة A تحتوي بالإجمالي حوالي 700 000 أسرة . لاحظنا على هذه العيّنة ، خلال شهر محدّد ، متوسّط استهلاك لهذه الأسر يساوي 950 ف ، بانحراف نموذجي يبلغ 600ف .

لنحسب فسحة الثقة العائدة إلى تقدير متوسّط استهلاك الأسر في المنطقة .

في هذا المثل :

$$N = 700\,000 , \qquad n = 10\,000$$
$$\bar{x} = 950 , \qquad s = 700 .$$

رغم كون العيّنة مسحوبة دون ردّ ، يمكننا عملياً ، بحكم ضعف نسبة البحث الإحصائي ، تشبيهها بعيّنة مستقلّة . في الواقع :

$$\frac{N-n}{N-1} = \frac{700\,000 - 10\,000}{699\,999} \neq 1 .$$

يتبع متوسّط العيّنة $\bar{x}$ قانوناً طبيعياً متوسّطه m ، وهو المتوسّط الحقيقي (المجهول) للمجتمع الإحصائي وانحرافه النموذجي $\sigma_{\bar{x}} = \sigma/\sqrt{n}$ ، حيث σ هو انحراف المجتمع الإحصائي النموذجي (مجهول) .

إذا كنّا نجهل قيمة σ الحقيقية ، نقدّرها انطلاقاً من العيّنة ، وبما أنّ مقدار العيّنة كبير:

$$s' \neq s = 700 .$$

الانحراف النموذجي $\sigma_{\bar{x}}$ لتوزيع متوسّط العيّنة يُقدّر إذن بواسطة :

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{700}{100} = 7 .$$

بحكم كبر حجم العيّنة ، فإنّ هذا التقدير دقيق بشكل كاف لأن يكون للمتغيّرة:

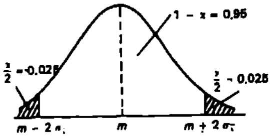
$$T = \frac{\bar{x} - m}{s/\sqrt{n}} = \frac{\bar{x} - m}{7} ,$$

توزيع طبيعي ممركز مختصر .

لنتفق مثلاً على درجة الاحتمال التالية : $1 - \alpha = 0.95$

بعبارة أخرى ، نقبل بمخاطرة باحتمال $\alpha = 0.05$ لأن نرتكب خطأ على دقّة التقدير . لنبحث عن القيمة t حيث :

$$P\{-t \leqslant T \leqslant +t\} = 0{,}95$$
$$P\{m - ts_{\bar{x}} \leqslant \bar{x} \leqslant m + ts_{\bar{x}}\} = 0{,}95 .$$



نقرأ في الجدول P(t) أو Π(t) :

$$t = 1{,}96 \neq 2 .$$

من هنا نستنتج فسحة الثقة ، بدرجة الاحتمال 95% ، المتماثلة بالنسبة للقيمة الملحوظة $\bar{x}$:

$$P\{\bar{x} - 2s_{\bar{x}} \leqslant m \leqslant \bar{x} + 2s_{\bar{x}}\} = 0{,}95 .$$

في الواقع ، إنّ عدم المساواة المزدوج :

$$m - ts_{\bar{x}} \leqslant \bar{x} \leqslant m + ts_{\bar{x}}$$

يعادل :

$$\bar{x} - ts_{\bar{x}} \leqslant m \leqslant \bar{x} + ts_{\bar{x}}$$

يوجد إذن 95 فرصة على 100 أن تكون قيمة متوسّط الاستهلاك الحقيقية m ضمن هذه الفسحة :

$$\bar{x} - 2s_{\bar{x}} \leqslant m \leqslant \bar{x} + 2s_{\bar{x}}$$
$$950 - (2 \times 7) \leqslant m \leqslant 950 + (2 \times 7)$$
$$936 \leqslant m \leqslant 964 .$$

كان يمكننا أن نظهر أكثر تصلّباً في ما يتعلّق بمخاطرة ارتكاب الخطأ على دقّة التقدير ونختار مثلاً درجة الاحتمال :

$$1 - \alpha = 0{,}99 .$$

قيمة t المناسبة التي نقرأها في الجدول P(t) أو Π(t) هي 2,58 . فسحة الثقة هي :

$$\bar{x} - 2{,}58\, s_{\bar{x}} \leqslant m \leqslant \bar{x} + 2{,}58\, s_{\bar{x}}$$
$$931{,}94 \leqslant m \leqslant 968{,}06 .$$

هنـاك 99 فرصـة على 100 لأن تنتمي قيمـة m الحقيقية إلى هـذه الفسحة . من الطبيعي أن تكون هذه الفسحة أكبر من سابقتها لأنّنا أردنا الحصول على فرص أقلّ في ارتكاب الخطأ . وإذا كنا نريد تصغير طول هذه الفسحة مع إبقائنا على نفس درجة الثقة لوجب علينا زيادة حجم العيّـنة .

مثل 2 . أجري بحث حول مجموع الرواتب الشهري x ، في مدينة صغيـرة ، بأخذ عيّنة تتألّـف من 50 موظّـفاً ، وكان معدّل البحث 1/10 . وقد حصلنا على النتائج الآتية :

$$\sum_i x_i = 75\,000 , \qquad \sum_i (x_i - \bar{x})^2 = 98\,000 .$$

حدّد فسحة الثقة بالنسبة لمتوسّـط الراتب ، بدرجة احتمال 95%. . في هذا المثل :

$$t = \frac{n}{N} = \frac{1}{10} , \qquad n = 50$$

$$\bar{x} = \frac{1}{n} \sum_i x_i = \frac{75\,000}{50} = 1\,500 \text{ F} .$$

متوسّـط العيّـنة $\bar{x}$ يتبع قانوناً طبيعياً متوسّـطه m وانحرافه النموذجي : $\sigma_{\bar{x}} = \sigma / \sqrt{n} \sqrt{(N-n)/(N-1)}$ لأنّ العيّـنة سُحِبَت دون رد .

انحراف المجتمع الإحصائي النموذجي مجهول ويتمّ تقديره بواسطة s' :

$$s'^2 = \frac{1}{n-1} \sum_i (x_i - \bar{x})^2 = \frac{98\,000}{49} = 2\,000$$

$$s' = \sqrt{2\,000} = 44{,}7 \text{ F} .$$

بالتالي ، نقدّر الانحراف النموذجي $\sigma_{\bar{x}}$ لتوزيع متوسّـط العيّـنة بواسطة $s_{\bar{x}}$:

$$s_{\bar{x}}^2 = \frac{s'^2}{n} \left(1 - \frac{n}{N}\right) = \frac{2\,000}{50} \left(1 - \frac{1}{10}\right) = 36$$

$$s_{\bar{x}} = 6 .$$

بافتراضنا أنّ للمتغيّـرة :

$$T = \frac{\bar{x} - m}{s_{\bar{x}}}$$

توزيعاً طبيعياً ممركزاً مختصراً ، لأنّ n أكبر من 30 ، نستنتج فسحة الثقة بدرجة 95% :

$$P\{\bar{x} - 2s_{\bar{x}} \leqslant m \leqslant \bar{x} + 2s_{\bar{x}}\} = 0{,}95 .$$

يوجد إذن 95 فرصة على 100 لأن تكون قيمة متوسّط الرواتب الحقيقية ضمن الفسحة :

$$\bar{x} - 2s_{\bar{x}} \leqslant m \leqslant \bar{x} + 2s_{\bar{x}}$$
$$1\,500 - 2 \times 6 \leqslant m \leqslant 1\,500 + 2 \times 6$$
$$1\,488 \leqslant m \leqslant 1\,512 .$$

ـ تقدير المجموع

لنفترض أنّه في المثل السابق أردنا تقدير مجموع كامل الرواتب ، وليس متوسّطها :

$$S = \sum_{s=1}^{N} X_s .$$

بناء على تعريف متوسّط المجتمع الإحصائي m :

$$S = \sum_{s=1}^{N} X_s = Nm .$$

ونقدّر مجموع الرواتب الكلّي بواسطة $N\bar{x}$ وانحرافه النموذجي هو $N\sigma_{\bar{x}}$ الذي نقدّره بدوره بواسطة $Ns_{\bar{x}}$.

بما أنّ N يساوي 500 ، فسحة الثقة بدرجة 95% هي :

$$N\bar{x} - 2Ns_{\bar{x}} \leqslant S \leqslant N\bar{x} + 2Ns_{\bar{x}}$$
$$750\,000 - 2 \times 500 \times 6 \leqslant S \leqslant 750\,000 + 2 \times 500 \times 6$$
$$744\,000 \leqslant S \leqslant 756\,000 .$$

B . تقدير النسبة

لنأخذ مجتمعاً إحصائياً مقداره N ، ويتألف من فئتين من الوحدات الإحصائية :

ـ الوحدات A بنسبة p ،

ـ الوحدات B بنسبة $q = 1 - p$

إنّ قيمة p مجهولة وننوي تقديرها بواسطة تردّد (تكرار) الوحدات A ، $f=x/n$ ، الملحوظ على العيّنة ذات الحجم n . هذا التردّد هو متغيّرة عشوائية يتوقّف قانون احتمالها على طريقة السحب ، مع أو بدون ردّ .

أ ـ عيّنة مستقلّة

بما أنّ سحب العيّنة تمّ مع ردّ ، فإنّه لا يغيّر النسبتين p وq .

عملياً ، نطابق العيّنة المسحوبة دون ردّ مع العيّنة المستقلّة عندما يكون مقدارها ضعيفاً بالنسبة لمقدار المجتمع الإحصائي N ، بشكل لا يؤثّر معه السحب على تكوين هذا المجتمع بشكل ملموس .

ضمن هذه السروط ، التردّد هو متغيّرة ذات حدّين (أنظر الفصل II ، ص 75) ، بمتغيّرين وسيطيّين n وp :

$$f = \mathscr{B}(n, p).$$

أملها الرياضي هو :

$$E\{f\} = p$$

وانحرافها النموذجي :

$$\sigma_f = \sqrt{\frac{pq}{n}}.$$

تسمح معرفة قانون احتمال f بتحديد فسحة ثقة التقدير عند درجة الاحتمال $1-\alpha$.

1 . عيّنة صغيرة

عندما يكون مقدار العيّنة n صغيراً جدّاً ، لا يمكننا أن نقرّب القانون ذا الحدّين من القانون الطبيعي أو من قانون بواسّون (Poisson) . وينبغي تحديد فسحة الثقة مباشرة انطلاقاً من القانون ذي الحدّين .

لكلّ قيمة ممكنة لِـ p ننسب قيمتين $f_1=x_1/n$ و$f_2=x_2/n$ بشكل يكون معه احتمال أن نشاهد f ضمن هذين الحدّين مساوياً تقريباً[1] لِـ $1-\alpha$:

$$\sum_{x \leqslant x_1} p(x) = \frac{\alpha}{2}, \qquad \sum_{x \geqslant x_2} p(x) = \frac{\alpha}{2},$$

حيث

$$p(x) = C_n^x\, p^x\, q^{n-x}.$$

(1) بما أنّ القانون ذا الحدّين منفصل (غير متّصل) ، لا يمكن بشكل عام إيجاد حدود تطابق تماماً الدرجة المعتمدة .

كون n و α ثابتين ، يمكننا أن نغيّر في قيم p ونحسب في كلّ حالة الحدّين f_1 وf_2 المناسبين ، وإذا نقلنا هذه القيم على رسم بياني ، نجد منحنيين C_1 وC_2 (الشكل 54) . بوسعنا إذاً أن نحدّد على الفور فسحة الثقة (p_1, p_2) التي تناسب التردّد $f=k/n$ الملحوظ على العيّنة .

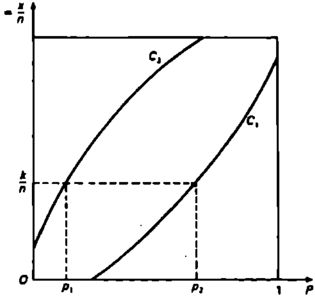
في الواقع ، لدينا تقريباً[1] :

$$\sum_{x \leq k} p_2(x) = \frac{\alpha}{2}, \qquad \sum_{x \geq k} p_1(x) = \frac{\alpha}{2}.$$

إذا كانت p أصغر من الحدّ الأدنى p_1 ، فإنّ احتمال أن نشاهد قيمة x تساوي k أو أكبر منها هو أصغر من $\alpha/2$. كذلك ، إذا كانت p أكبر من الحدّ الأعلى p_2 $(p>p_2)$ ، فإنّ احتمال أن نشاهد قيمة x تساوي k أو أصغر منها هو أصغر من $\alpha/2$. الحاصل ، هناك إذن احتمال $1-\alpha$ لأن تكون قيمة p الحقيقية ضمن الفسحة (p_1, p_2) .

لقد تمّ وضع لوحات بيانية (مع منحنيات) حسب نموذج الشكل 54 ، وهي تقدّم بالنسبة لدرجة احتمال محدّدة ، شبكة المنحنيات C_1 وC_2 التي تناسب مختلف قيم n . ونجد في الملحق اللوحة البيانية 1 التي تناسب درجة الاحتمال $1-\alpha = 95\%$.

أمّا قيم p_1 وp_2 العددية التي تطابق هذه اللوحات البيانية فنجدها في جداول فيشر ويايتس (Fisher and Yates)[2]



الشكل 54 . فسحة الثقة (p_1, p_2) : المناسبة للتردّد k/n الملحوظ على العيّنة

(1) انظر الملاحظة السابقة .

(2) R.A. Fisher and yates, «Statistical tables for biological, agricultural and medical research», London, Oliver and Boyd, 1963.

مثلاً : لقد أخذنا من كمّية من القطع المصنوعة من مادّة لدائنية معيّنة عيّنة تتألّف من 10 قطع ظهر منها 3 معيبة عند الفحص .

لنفترض أن العيّنة سُحبت مع ردّ أو أنّ مقدار الكمّية كبير بشكل كاف لجعل السحب لا يؤثّر ، عملياً ، على تكوين هذه الكمّية .

في هذا المثل :

$$n = 10, k = 3, f = k/n = 0{,}3$$

لنحدّد درجة الاحتمال $1-\alpha$ ، مثلاً بـ 95% .

نحدّد فسحة الثقة بواسطة الحدّين p_1 وp_2 حيث :

$$\sum_{x \geq k} p_1(x) = \frac{\alpha}{2} \qquad \sum_{x \leq k} p_2(x) = \frac{\alpha}{2}.$$

نحصل هكذا على المعادلتين التاليتين :

$$\sum_{x=3}^{10} C_{10}^{x} p_1^{x} q_1^{10-x} = C_{10}^{3} p_1^{3} q^{7} + C_{10}^{4} p_1^{4} q^{6} + \cdots + C_{10}^{10} p_1^{10} = 0{,}025 \qquad (1)$$

$$\sum_{x=0}^{3} C_{10}^{x} p_2^{x} q_2^{10-x} = C_{10}^{0} q_2^{10} + C_{10}^{1} p_2 q_2^{9} + \cdots + C_{10}^{3} p_2^{3} q_2^{7} = 0{,}025. \qquad (2)$$

إذا أخذنا المتمّم إلى 1 من عنصري المعادلة (2) ، تصبح مطابقة لـ :

$$\sum_{x=4}^{10} C_{10}^{x} p_2^{x} q_2^{10-x} = 0{,}975. \qquad (3)$$

يمكننا حلّ المعادلتين (1) و(3) من خلال جداول القانون ذي الحدّين التي سبق ذكرها (أنظر الفصل II ، ص 78) ، والتي تعطينا قيم :

$$\sum_{x=k}^{n} C_n^x p^x q^{n-x} \quad \text{بالنسبة لـ} \quad n = 1, 2, \ldots, 100.$$

وقد تمّ إجراء هذه الحلول ووضعها بشكل نهائي في جدول فيشر ويتس ، الذي يعطي مباشرة قيمتي p_1 وp_2 المرجوّتين . من جهة أخرى ، الطريقة الأسهل هي مراجعة اللوحة البيانية (الملحق : اللوحة البيانية 1) . فنجد

$$p_1 = 0{,}07, \; p_2 = 0{,}65$$

هناك إذن تقريباً 95 فرصة على 100 كي تكون نسبة القطع المعيبة الحقيقية موجودة

ضمن الفسحة :

$$0,07 \leqslant p \leqslant 0,65$$

كما نلاحظ ، يستدعي تحديد فسحة الثقة من خلال القانون ذي الحدّين إمّا إجراء حسابات شاقّة بعض الشيء ، إمّا مراجعة وثائق غير متداولة كثيراً (جداول القانون ذي الحدّين ، جداول فيشر ويتس ، لوحات بيانية) . لحسن الحظ ، ما أن يكون مقدار العيّنة كبيراً بما فيه الكفاية ، يصبح تقريب القانون ذي الحدّين من قانون بواسّون أو من القانون الطبيعي (المعتدل) ممكناً ، ممّا يسهّل الحسابات كثيراً .

2 . التقريب من قانون بواسّون

عندما تكون n كبيرة وp صغيرة ، بشكل يبقى معه حاصل الضرب np مساوياً لبضعة آحاد ، يمكن تقريب القانون ذي الحدّين من قانون بواسّون بمتغيّر وسيطي m =np (أنظر الفصل II ، ص 94) . عملياً ، نعتبر التقريب صالحاً عندما يكون لدينا في آن واحد :

$$p < 0,10 \text{ و } n > 50$$

ويجري تحديد فسحة الثقة تبعاً لنفس المبادىء السابقة لكن الحسابات أبسط والاستعمال الممكن للجداول أو اللوحات البيانية أسهل كون قانون بواسّون لا يتعلّق سوى بمتغيّر وسيطي واحد ، بدلاً من متغيّرين اثنين n وp ، بالنسبة للقانون ذي الحدّين .

عند درجة الثقة $1-\alpha$ ، نبحث عن القيمتين p_1 وp_2 حيث ، تقريباً[1]

$$\sum_{x \geqslant k} p_1(x) = \frac{\alpha}{2}, \qquad \sum_{x \leqslant k} p_2(x) = \frac{\alpha}{2}.$$

k هو عدد الوحدات A الملحوظ على العيّنة .

في هاتين المعادلتين :

$$m_1 = np_1, \quad \text{مع} \quad p_1(x) = \frac{e^{-m_1} m_1^x}{x!}$$

$$m_2 = np_2, \quad \text{مع} \quad p_2(x) = \frac{e^{-m_2} m_2^x}{x!}$$

(1) قانون بواسّون هو ، كالقانون ذي الحدّين ، منفصل ، وليس من الممكن بشكل عام إيجاد حدود تطابق تماماً الدرجة المعتمدة .

إذا كانت النسبة p أصغر من p_1 $(p< p_2)$ ، فإنّ احتمال أن نشاهد قيمة x أكبر من أو تساوي k هو أصغر من $\alpha/2$. كذلك ، إذا كانت النسبة p أكبر من p_2 $(p>p_2)$ ، فإنّ احتمال أن نشاهد قيمة x أصغر من أو تساوي k هو أصغر من $\alpha/2$. هناك إذن الاحتمال $1-\alpha$.أن توجد قيمة p الحقيقية ضمن الفسحة (p_1 , p_2) .

ونقوم بحلّ هاتين المعادلتين باستعمال جدول قانون بواسّون أو ، أفضل ، بمراجعة لوحة بيانية وضِعت بشكل مماثل للوحة القانون ذي الحدّين . ونجد في الملحق اللوحة البيانية 2 التي تناسب درجة الاحتمال $1 - \alpha = 95\%$

مثلاً : في إحدى الصيدليات ، تحتوي كمّية البضاعة على عشرة آلاف سلعة مختلفة وتجري عملية الجرد مرّة في السنة . كي نفحص دقّة هذه العملية ، سحبنا عيّنة تتألّف من 100 سلعة ، ووجدنا 4 أخطاء في كشف حسابها .

لدينا في هذا المثل :

$$n = 100, k = 4, f = k/n = 0,04$$

لقد اجتمعت شروط تطبيق قانون بواسّون : مقدار العيّنة n كبير بدرجة كافية و p ، التي نقدّرها بواسطة f ، هي نسبة مئوية ، بشكل يساوي معه حاصل الضرب np بضعة آحاد .

لنحدّد درجة الاحتمال $1-\alpha$ ، مثلاً 95%

كي نجد فسحة الثقة ، يكفي أن نبحث في جدول قانون بواسّون عن القيمتين m_1 و m_2 حيث ، تقريباً :

$$\sum_{x \geq 4} \frac{e^{-m_1} m_1^x}{x!} = 0,025 \qquad (1)$$

$$\sum_{x \leq 4} \frac{e^{-m_2} m_2^x}{x!} = 0,025 . \qquad (2)$$

إذا أخذنا المتمّم إلى 1 من كلّ من عنصري المعادلة (1) ، فإنها تطابق :

$$\sum_{x<4} \frac{e^{-m_1} m_1^x}{x!} = 0,975 \qquad (3)$$

ما يلائم مراجعة الجدول (الملحق : الجدول 1) . حيث نقرأ :

$$\sum_{x<4} \frac{e^{-1} 1^x}{x!} = 0.981\,0 \neq 0.975$$

$$\sum_{x \leq 4} \frac{e^{-10}\, 10^x}{x\,!} = 0{,}029\,3 \neq 0{,}025\,.$$

لدينا إذن :

$$m_1 = np_1 = 1\,, \qquad m_2 = np_2 = 10\,.$$

ونستنتج فسحة ثقة p عند درجة الاحتمال $1-\alpha = 0{,}95$:

$$0{,}01 \leq p \leq 0{,}10$$

وهذه هي بالفعل النتيجة التي يمكننا قراءتها على اللوحة البيانية 2 عند k=4 .

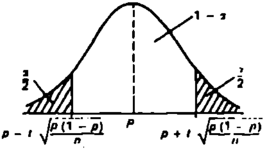
لو كنّا نرغب بدقّة أكبر في ما يتعلّق بحدّي هذه الفسحة p_1 وp_2 ، لكان ينبغي استعمال جدول قانون بواسّون حيث يتغيّر المتغيّر الوسيطي m كلّ عِشر (من عشر إلى عِشر) (أنظر الفصل II ، ص 97) .

3 . التقريب من القانون الطبيعي (المعتدل)

عندما يكون مقدار العيّنة n كبيراً دون أن تتحقّق شروط تطبيق قانون بواسّون ـ p ليست قريبة من صفر ولا من 1 ـ يمكننا تقريب القانون ذي الحدّين الذي يتبعه التردّد f الملحوظ على العيّنة من قانون طبيعي . عادةً ، نعتبر تقريب القانون ذي الحدّين من القانون الطبيعي صحيحاً عندما يتجاوز كلّ من np وnq من 15 إلى 20 وحدة . إذا كانت العيّنة مستقلّة ، أو يمكنها ، على الأقلّ أن تُعتبر كذلك (يكون مقدار العيّنة n ضعيفاً بالنسبة لمقدار المجتمع الإحصائي N) فإنّ متغيّري هذا القانون الوسيطيّين هما :

$$m = p \qquad و \qquad \sigma = \sqrt{\frac{p(1-p)}{n}}\,.$$

كلّما كان استعمال القانون الطبيعي ممكناً فإنّه يسهّل إلى حدّ بعيد تحديد فسحة الثقة .



لنختر درجة احتمال $1-\alpha$.

لهذه الدرجة تناسب قيمة معيّنة t للمتغيّرة الطبيعية المركزة المختصرة التالية

$$T = \frac{f - p}{\sqrt{p(1-p)/n}}$$

حيث :

$$P\{-t \leqslant T \leqslant +t\} = 1-\alpha$$

$$P\left\{ p - t\sqrt{\frac{p(1-p)}{n}} \leqslant f \leqslant p + t\sqrt{\frac{p(1-p)}{n}} \right\} = 1-\alpha .$$

من هنا نستنتج فسحة الثقة ، عند درجة الاحتمال $1-\alpha$:

$$P\left\{ f - t\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{p(1-p)}{n}} \right\} = 1-\alpha$$

في الواقع ، إنّ عدم المساواة المزدوج :

$$p - t\sqrt{\frac{p(1-p)}{n}} \leqslant f \leqslant p + t\sqrt{\frac{p(1-p)}{n}}$$

يعادل :

$$f - t\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{p(1-p)}{n}}$$

إلاّ أنّـه بما أنّ قيمة p مجهولة فإنّنا لا نعرف قيمة $\sqrt{p(1-p)/n}$.
يمكننا اعتماد طريقتين لحلّ هذه المشكلة .

طريقة القَطع الإهليلجي (ellipse)

إنّ عدم المساواة التالي :

$$f - t\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{p(1-p)}{n}}$$

$$-t\sqrt{\frac{p(1-p)}{n}} \leqslant p - f \leqslant +t\sqrt{\frac{p(1-p)}{n}}$$

يعادل :

$$|p-f| \leqslant t\sqrt{\frac{p(1-p)}{n}} .$$

إذا رفعنا عنصري عدم المساواة هذه ، وهما إيجابيان ، إلى مربّعيهما ، نحصل على :

$$(p-f)^2 \leqslant t^2\frac{p(1-p)}{n} .$$

ما يمكننا كتابته :

$$(p-f)^2 - t^2\frac{p(1-p)}{n} \leqslant 0$$

أي ، إذا وسّعنا :

$$p^2\left(1+\frac{t^2}{n}\right)-p\left(2f+\frac{t^2}{n}\right)+f^2\leqslant 0. \qquad (1)$$

ويعطينا حلّ هذه المباينة حدّي فسحة الثقة p_1 وp_2 التي تناسب درجة الاحتمال $1-\alpha$. عند قيمة محدّدة لمقدار العيّنة n ولـ t ، قيمة المتغيّرة الطبيعية المركزة المختصرة والتي تناسب درجة الإحتمال $1-\alpha$ ، المعادلة

$$p^2\left(1+\frac{t^2}{n}\right)-p\left(2f+\frac{t^2}{n}\right)+f^2=0$$

هي معادلة قطع إهليلجي (ellipse) في السطح (p, f) . ويتحقّق عدم المساواة (1) عند النقاط الموجودة داخل هذا القطع الإهليلجي .

من ناحية أخرى ، لدينا حتماً :

$$0\leqslant p\leqslant 1\,;\,0\leqslant f\leqslant 1$$

إذن يجب الأخذ بعين الاعتبار فقط أجزاء القطع الإهليلجي التي تناسب قسم السطح المحدّد بهذه المباينات . هكذا نحصل على رسم بياني يتضمّن قوسين من القطع الإهليلجي C_1 وC_2 ، كثير الشبه بالشكل 54 . ويسمح لنا هذا الرسم البياني بإيجاد فسحة الثقة (p_1 , p_2) التي تناسب التردّد $f = k/n$ الملحوظ على العيّنة على الفور .

حسب هذا النموذج ، تمّ وضع لوحات بيانية تقدّم ، بالنسبة لدرجة احتمال معيّنة ، شبكة المنحنيات C_1 وC_2 التي تناسب مختلف قيم n . وتجمع هذه اللوحات على نفس الرسم البياني أقواس المنحنيات المأخوذة انطلاقاً من القانون ذي الحدّين إذا كانت $n\leqslant 100$ ، وأقواس القطع الإهليلجية المحدّدة بواسطة التقريب من القانون الطبيعي إذا كانت $n>100$. وهذا حال اللوحة البيانية 1 ، التي سبق ذكرها ، والتي نجدها في الملحق والتي تناسب درجة الاحتمال $1-\alpha = 95\%$.

هذه الطريقة ، الشبيهة بالطريقة المستعملة في حالة القانون ذي الحدّين وقانون بواسّون ، هي الطريقة الدقيقة الوحيدة . وعندما لا تكون اللوحات البيانية بتصرّفنا ، يكون بوسعنا ، حسب النهج المعروض لاحقاً ، الحصول على تقريب جيّد لفسحة الثقة .

طريقة تقدير الانحراف النموذجي

نحدّد فسحة ثقة تقدير p بواسطة :

$$f - t\sqrt{\frac{p(1-p)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{p(1-p)}{n}} \cdot$$

بما أنّ f هي مقدّر غير متحيّز لـ p ، يمكننا تقدير الانحراف النموذجي $\sigma_f = \sqrt{p(1-p)/n}$ بواسطة $\sqrt{f(1-f)/n}$.[1] . ويُسمح بهذا الاستبدال لأنّه ، كون n كبيرة ، فإنّ المتغيّرة الطبيعية المركزة المختصرة :

$$T = \frac{f - p}{\sqrt{f(1-f)/n}}$$

يمكن اعتبارها موزّعة تقريباً حسب القانون الطبيعي .

نأخذ إذن فسحة الثقة :

$$f - t\sqrt{\frac{f(1-f)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{f(1-f)}{n}}$$

ويكون لدينا تقريباً :

$$P\left\{f - t\sqrt{\frac{f(1-f)}{n}} \leqslant p \leqslant f + t\sqrt{\frac{f(1-f)}{n}}\right\} = 1 - \alpha .$$

فما أن يكون مقدار العيّنة n كبيراً بما فيه الكفاية ، متجاوزاً المائة وحدة ، فإنّ التقريب الناتج عن هذه الطريقة يصبح جيّداً جدّاً .

مثل 1 . في مجمّع سكّاني كبير ، جرى بحث إحصائي لتحديد نسبة الأسر التي تملك سيارة . سحبنا عيّنة مستقلّة تتكوّن من 2000 أسرة ووجدنا بينها 600 مالكة لسيّارة واحدة على الأقلّ .

في هذا المثل :

$$n = 2\,000 , \qquad f = \frac{600}{2\,000} = 0{,}3 .$$

---

(1) بشكل أدقّ $\sqrt{f(1-f)/(n-1)}$ هو مقدّر غير متحيّز لـ $\sigma_f = \sqrt{p(1-p)/n}$: انظر الفقرة 1.b ، ص 251 .

يتوزّع التكرار f حسب قانون ذي حدّين بمتغيّرين وسيطيّين n ، مقدار العيّنة ، وp ، النسبة المجهولة للأسر التي تملك سيارة . بما أنّ n كبيرة وp ليست قريبة من صفر ولا من 1 ، يمكننا تقريب هذا القانون من توزيع طبيعي متوسّطه p وانحرافه النموذجي $\sqrt{p(1-p)/n}$.

طريقة تقدير الانحراف النموذجي

بما أنّنا نجهل قيمة p الحقيقية ، فإنّنا نقدّر الانحراف النموذجي بواسطة :

$$\sqrt{\frac{f(1-f)}{n}}$$

قيمة نحصل عليها باستبدالنا p بالتردّد f الملحوظ على العيّنة ، وهذا الاستبدال ممكن بحكم حجم العيّنة المرتفع .

لنحدّد درجة الاحتمال ، مثلاً : $1-\alpha = 0{,}95$ ولنبحث في جدولي القانون الطبيعي p(t) أو π(t) عن قيمة t حيث :

$$P(-t \leqslant T \leqslant +t) = 0{,}95$$

$$P\left\{p - t\sqrt{\frac{f(1-f)}{n}} \leqslant f \leqslant p + t\sqrt{\frac{f(1-f)}{n}}\right\} = 0{,}95 .$$

فنجد ، كما نعرف :

$$t = 1{,}96 .$$

من هنا نستنتج فسحة ثقة تقدير p ، عند درجة الاحتمال 95% :

$$P\left\{f - 1{,}96\sqrt{\frac{f(1-f)}{n}} \leqslant p \leqslant f + 1{,}96\sqrt{\frac{f(1-f)}{n}}\right\} = 0{,}95 .$$

بالنسبة لدرجة الاحتمال هذه ، غالباً ما نكتفي ، للسهولة ، بحساب فسحة الثقة بواسطة القيمة القريبة t=2 . هنا نستعمل قيمة t الحقيقية للحصول على فسحة ثقة دقيقة كي يمكن مقارنتها مع الفسحة المحسوبة بواسطة الطريقة الأدقّ ، طريقة القطع الإهليلجي .

لدينا :

$$\sqrt{\frac{f(1-f)}{n}} = \sqrt{\frac{0,3 \times 0,7}{2\,000}} = 0,010\,2 .$$

إذا نقلنا هذه القيمة في عبارة الفسحة نجد :

$$f - 1,96 \sqrt{\frac{f(1-f)}{n}} \leqslant p \leqslant f + 1,96 \sqrt{\frac{f(1-f)}{n}}$$
$$0,3 - 1,96 \times 0,010\,2 \leqslant p \leqslant 0,3 + 1,96 \times 0,010\,2$$
$$0,280\,0 \leqslant p \leqslant 0,320\,0 .$$

طريقة القطع الإهليلجي

إنّ حلّ عدم المساواة :

$$|p - f| \leqslant t \sqrt{\frac{p(1-p)}{n}}$$

يعني حلّ عدم المساواة التالي ، وهو من الدرجة الثانية حسب p :

$$p^2\left(1 + \frac{t^2}{n}\right) - p\left(2f + \frac{t^2}{n}\right) + f^2 \leqslant 0$$

وإذا وضعنا t و n و f بقيمها :

$$t = 1,96 \qquad n = 2000 \qquad f = 0,3$$

نحصل على :

$$1,0019p^2 - 0,601\,p + 0,0900 \leqslant 0$$

جذرا معادلة الدرجة الثانية المناسبة هما :

$$p_1 = 0,2804 \qquad p_2 = 0,3203$$

وهما حدّا فسحة الثقة :

$$0,2804 \leqslant p \leqslant 0,3203$$

هذه النتيجة هي معادلة للنتيجة التي وجدناها أعلاه . آخذين بعين الاعتبار الدقّة المرجوّة في هذا النوع من المعلومات ، يكفي في الواقع أن نستطيع التأكيد على وجود 95 فرصة من 100 أن تكون القيمة الحقيقية لنسبة الأسر التي تملك سيارة موجودة في الفسحة :

$$0,28 \leqslant p \leqslant 0,32$$

إذن عندما يكون مقدار العيّنة مرتفعاً بما فيه الكفاية ، لا نتردّد في حساب فسحة الثقة مقدّرين الانحراف النموذجي $\sqrt{p(1-p)/n}$ بواسطة $\sqrt{f(1-f)/n}$.

ويمكن قراءة هذه النتائج مباشرة على اللوحة البيانية 1 .

مثل 2 . في مدينة معيّنة جرى بحث إحصائي على عيّنة مستقلّة تتضمّن 586 أسرة لمعرفة ما إذا كانت راضية أو غير راضية عن شروط سكنها : وقد صرّح 57% من الأسر عن رضاهم .

في هذا المثل :

$$n = 586, \qquad f = 0,57$$

لقد تحقّقت شروط تقريب القانون ذي الحدّين الذي تتبعه f ، من قانون طبيعي متوسّطه p ، النسبة الحقيقية للأسر الراضية ، وانحرافه النموذجي $\sqrt{p(1-p)/n}$.

لنختر درجة الاحتمال :

$$1 - \alpha = 0,95$$

التي تناسبها :

$$t \# 2 .$$

لدينا :

$$P\left\{ f - 2\sqrt{\frac{p(1-p)}{n}} < p < f + 2\sqrt{\frac{p(1-p)}{n}} \right\} = 0,95 .$$

إذا وضعنا القيمة الملحوظة f مكان p في عبارة الانحراف النموذجي :

$$\sqrt{\frac{f(1-f)}{n}} = \sqrt{\frac{0,57 \times 0,43}{586}} = 0,020$$

نحصل على فسحة الثقة :

$$f - 2\sqrt{\frac{f(1-f)}{n}} < p < f + 2\sqrt{\frac{f(1-f)}{n}}$$

$$0,57 - 2 \times 0,02 < p < 0,57 + 2 \times 0,02$$

$$0,53 < p < 0,61 .$$

ب ـ عيّنة مستنفدة

عندما نجري السحب دون ردّ ، فإنّ عدد الوحدات A ، x الملحوظ على العيّنة يتبع قانوناً فوق هندسي . وأمل التردّد $f = x/n$ الرياضي هو :

$$E\{f\} = p$$

وتباينه :

$$V\{f\} = \frac{p(1-p)}{n}\cdot\frac{N-n}{N-1} \approx \frac{p(1-p)}{n}\left(1-\frac{n}{N}\right).$$

لتحديد فسحة الثقة ، نتّبع نفس طريقة التفكير كما في حالة العيّنة المستقلّة ، لكن الحسابات معقّدة أكثر لأنّنا نستبدل القانون ذا الحدّين بالقانون فوق الهندسي .

ويمكن إجراء حسابات تقريبية في حالتين تصادفان كثيراً لحسن الحظ .

1 . التقريب من القانون ذي الحدّين

كما سبق أن أشرنا ، عندما يكون مقدار العيّنة n صغيراً بالنسبة لمقدار المجتمع الإحصائي ، بشكل لا يؤثّر فيه السحب على تكوين هذا المجتمع بشكل ملموس (عملياً تكون نسبة البحث n/N أصغر من 10%) ، يمكننا تشبيه العيّنة المستنفدة بعيّنة مستقلّة . في الواقع ، ضمن هذه الشروط يمكننا تقريب القانون فوق الهندسي من القانون ذي الحدّين (أنظر الفصل II ، ص 87) ، الذي يمكننا استبداله بدوره ، حسب قيم n وp ، بقانون بواسّون متغيّره الوسيطي m=np أو بقانون طبيعي متوسّطه p وانحرافه النموذجي $\sqrt{p(1-p)/n}$.

من جهة أخرى سوف نلاحظ أنّه حتى في حال عدم تحقّق شرط التقريب من القانون ذي الحدّين فإنّ استعماله يعطينا ، عملياً ، تقديراً نحو الزيادة لفسحة الثقة .

2 . التقريب من القانون الطبيعي

عندما يكون في الوقت نفسه مقدار المجتمع الإحصائي N ومقدار العيّنة n كبيرين ، ولا يمكن إهمال n بالنسبة لِـ N ، فإنّ القانون فوق الهندسي الذي يتبعه التردّد f يمكن تقريبه من قانون طبيعي أمله الرياضي :

$$E\{f\} = p$$

وانحرافه النموذجي :

$$\sigma_f = \sqrt{\frac{p(1-p)}{n}\cdot\frac{N-n}{N-1}}.$$

هذا الميل للقانون فوق الهندسي نحو القانون الطبيعي ينتج عن ما سبق أن عرضناه في ما يخصّ قانون توزيع متوسّط عيّنة كبيرة : يمكننا في الواقع اعتبار التردّد f كمتوسّط

n متغيّرة برنولي غير مستقلّة متبار التردّد f كمتوسّط (أنظر الفصل III ، ص 114) .

ينبغي أن لا ننسى العامل التصحيحي :

$$\frac{N-n}{N-1} \neq 1-\frac{n}{N},$$

الذي يُسمّى أحياناً مُعامِل الاستنفاد والذي يصغّر فسحة الثقة كلّما مالت نسبة البحث الإحصائي t = n/N نحو 1 . أقصى ما يمكن ، يصبح مقدار العيّنة مساوياً لمقدار المجتمع الإحصائي : تتمّ ملاحظة كلّ الوحدات الإحصائية . لا يعود التردّد f متغيّرة عشوائية ، إنّها تساوي عندئذٍ p والانحراف النموذجي يساوي صفراً .

مثلاً . غالباً ما تنتج الإحصاءات المستخدمة لوضع لوحة قيادة شركة معيّنة عن استعمال عدد من الوثائق الأساسية ، وتأخذ هذه العمليات فترة معيّنة . ويسمح لنا استعمال هذه الوثائق عن طريق البحث الإحصائي بوضع هذه المعلومات بسرعة في تصرّف المسؤولين ، بدقّة مقبولة تماماً .

في مشروع تجاري معيّن ، تمّ تسجيل 4230 تعليمة خلال فترة محدّدة . وجرى استخدام سريع لهذه الوثائق على عيّنة بمقدار الـ 1/5 مسحوبة دون ردّ : استنتجنا أنّ 119 تعليمة (طلباً) لم تُلبّى .

في هذا المثل :

$$N = 4230 ; n = 4230/5 = 846 ; f = 119/846 = 0,141$$

لنختر درجة الاحتمال : $1-\alpha = 0,95$ التي تتناسب مع : $t \neq 2$.

لدينا :

$$P\left\{f-2\sqrt{\frac{p(1-p)}{n}}\left(1-\frac{n}{N}\right) \leqslant p \leqslant f+2\sqrt{\frac{p(1-p)}{n}}\left(1-\frac{n}{N}\right)\right\} = 0,95 .$$

إذا استبدلنا p في عبارة الانحراف النموذجي بالقيمة الملحوظة f :

$$\sqrt{\frac{f(1-f)}{n}}\left(1-\frac{n}{N}\right) = \sqrt{\frac{0.141 \times 0.859}{846}}\left(1-\frac{1}{5}\right) = 0.011 .$$

نحصل على فسحة الثقة :

$$0,141 - 2 \times 0,011 \leq p \leq 0,141 + 2 \times 0,011$$
$$0,118 \leq p \leq 0,163$$

يوجد 95 فرصة على 100 أن تكون نسبة الطلبات التي لم تُلبّى خلال هذه الفترة محصورة بين 11,8 و16,3% .

ـ تقدير المقدار

لنعد إلى المثل 1 ، ص 267 ولنفترض أنّ عدد الأسر الموجودة في التجمّع السكني يبلغ $N = 80\ 000$. هذه المرّة ننوي تقدير ، ليس النسبة p للأسر التي تملك سيّارة ، بل عدد هذه الأسر N_A :

$$N_A = Np$$

يتمّ تقدير هذا العدد بواسطة Nf .

ونستنتج فسحة ثقة هذا التقدير تلقائياً من الفسحة العائدة إلى تقدير p :

$$p_1 \leq p \leq p_2$$
$$Np_1 \leq N_A \leq Np_2$$

عند درجة الاحتمال $1-\alpha = 0,95$ ، كان لدينا :

$$0,28 \leq p \leq 0,32$$

بالتالي :

$$0,28 \times 80\ 000 \leq N_A \leq 0,32 \times 80\ 000$$
$$22\ 400 \leq N_A \leq 25\ 600$$

C . تحديد حجم العيّنة

يعلّمنا قانون الأعداد الكبيرة أنّه يكفي سحب عيّنة بمقدار كاف للحصول بصفة شبه مؤكّدة على الدقّة المطلوبة لتقدير متغيّر وسيطي لمجتمع إحصائي معيّن .

المسألة التي تطرح نفسها هي إذن التالية : بإعطائنا مسبقاً درجة احتمال $1-\alpha$ معيّنة ، كم يجب أن يكون مقدار العيّنة للحصول على تقدير بالدقّة المطلوبة ؟

أ ـ تقدير المتوسّط

يمكننا اعتبار توزيع متوسّط عيّنة كبيرة $\bar{x}$ توزيعاً طبيعياً أمله الرياضي m وانحرافه النموذجي :

في حالة عيّنة مستقلة : $\frac{\sigma}{\sqrt{n}}$

في حالة عيّنة مستنفدة : $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

لنفترض أن سحب العيّنة هو سحب مع ردّ أو أنّه يمكننا اعتباره كذلك .
تتناسب درجة الاحتمال $1-\alpha$ مع فسحة الثقة التالية :

$$\bar{x} - t_\alpha \frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \bar{x} + t_\alpha \frac{\sigma}{\sqrt{n}}$$

أي :

$$|\bar{x} - m| \leqslant t_\alpha \frac{\sigma}{\sqrt{n}}.$$

كي تكون دقّة التقدير تساوي على الأقلّ k% من m (دقّة محدّدة بالقيمة غير المطلقة) ، يجب اختيار n بشكل :

$$t_\alpha \frac{\sigma}{\sqrt{n}} \leqslant km$$

أي

$$\sqrt{n} \geqslant \frac{t_\alpha}{k}\frac{\sigma}{m}, \qquad n \geqslant \frac{t_\alpha^2}{k^2}\frac{\sigma^2}{m^2}.$$

في العنصر الثاني من هذه المباينة نتعرّف إلى عبارة مُعامل التغيّر (أنظر «الإحصاء الوصفي»، الفصل VI ، القسم II ، الفقرة 4.D) :

$$CV = \frac{\sigma}{m},$$

الذي يقيس تشتّت المتغيّرة X النسبي .
علينا إذن أن نختار :

$$n \geqslant \frac{t_\alpha^2}{k^2}(CV)^2 .$$

تُظهر هذه العبارة أنّ حجم العيّنة ، عند درجة احتمال ودقّة معيّنتين ، هو قيمة تناسبية مع مربّع معامل التغيّر : هو أضعف بالنسبة لمجتمع إحصائي قليل التشتّت منه

بالنسبة لمجتمع إحصائي مشتّت جدّاً

كي نحدد حجم العيّنة ينبغي إذن معرفة القيمة $CV = \sigma/m$.

. لكن كوننا نجهل قيمة m التي نبحث بالضبط عن تقديرها ، فإنّنا نجهل بطبيعة الحال قيمة σ/m التي تدخل الانحراف النموذجي . إلاّ أنّه في عدد من الحالات لا يكون معامل التغيّر مجهولاً تماماً ، ومعرفته ، حتّى على وجه التقريب ، الناتجة مثلاً عن بحث إحصائي سابق ، تسمح باختيار قيمة معقولة لـ n . وبعد ذلك ، يمكننا حساب الدقّة الحاصلة حقيقة .

بالمقابل ، إذا لم يكن لدينا اي فكرة عن قيمة σ/m ، لا يمكننا أن نحلّ المسألة المطروحة ، ونضطر عندها إلى إجراء البحث الإحصائي على مرحلتين : تخدمنا المرحلة الأولى ، التي نجريها على عيّنة محدودة ، في تقييم مُعامل التغيّر ، ونحدّد للمرحلة الثانية حجم العيّنة النهائي .

مثلاً . في مجتمع إحصائي معيّن يبلغ مُعامل تغيّر ما يُنفق على مستحضرات الزينة تقريباً 4 . حدّد حجم العيّنة الذي يخوّلنا تقدير قيمة متوسّط هذه النفقة بدقّة 10% وبدرجة احتمال $1 - \alpha = 0,95$.

في هذا المثل :

$$\frac{\sigma}{m} = 4 , \qquad k = 0,10 .$$

تناسب درجة الاحتمال :

$$1 - \alpha = 0,95$$

مع القيمة : $t \neq 2$ ، من قيم المتغيّرة الطبيعية المركزة المختصرة .
يجب إذن أن نختار :

$$n \geqslant \frac{2^2}{(0,1)^2} \times 16 = 6\,400 .$$

يمكننا بسهولة أن نبسط هذا الاستدلال مثلاً إلى الحالة حيث لا يمكننا تشبيه سحب العيّنة بسحب مع ردّ وحيث تحدّد الدقّة المطلوبة بالقيمة المطلقة .

مثلاً . يتمّ تسليم قساطل (أنابيب) مصنوعة بالجملة من مادة بلاستيكية على

. كمّيات تتضمّن كلّ منها 200 . بناء على طلب معيّن ، قُرّر بالنسبة لكلّ كمّية تقدير متوسّط طول الأنابيب بواسطة البحث الإحصائي . مع العلم أن الانحراف النموذجي لتوزيع طول هذه الأنابيب يبلغ 4 ملم ، حدّد حجم العيّنة التي يجب فحصها في كلّ كمّية كي يكون الخطأ على تقدير متوسّط الطول ، بالنسبة لـ 95 كمّية على 100 ، أصغر من 0,80 ملم .

يتمّ سحب العيّنة دون ردّ وحجم الكمّية لا يكفي لتشبيه طريقة السحب هذه بسحب عيّنة مستقلّة .

تتناسب درجة الاحتمال $1-\alpha$ مع فسحة الثقة :

$$|\bar{x} - m| < t_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}}.$$

كي تكون دقّة التقدير تساوي على الأقلّ a (دقّة محدّدة بالقيمة المطلقة) ، يجب اختيار n بالشكل :

$$t_\alpha \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N - n}{N - 1}} \leqslant a$$

أي :

$$n \geqslant \frac{t_\alpha^2 \sigma^2 N}{t^2 \sigma^2 + a^2(N - 1)}.$$

في هذا المثل :

$$N = 200, \quad \sigma = 4 \text{ mm}, \quad a = 0{,}80 \text{ mm}.$$

تتناسب درجة الاحتمال :

$$1 - \alpha = 0{,}95$$

مع :

$$t \neq 2.$$

لدينا إذن :

$$2 \frac{4}{\sqrt{n}} \sqrt{\frac{200 - n}{200 - 1}} \leqslant 0{,}80$$

أي ، إذا اختزلنا ورفعنا عنصري عدم المساواة إلى مربّعيها :

$$\frac{1}{n}\frac{200-n}{199} \leqslant 0{,}01$$

$$n \geqslant \frac{200}{2{,}99} = 67 .$$

كي نحصل على الدقّة المطلوبة علينا إذن أن نقيس في كلّ كمّية طول 67 أنبوباً نسحبها بالصدفة .

ب ـ تقدير النسبة

يمكننا اعتبار النسبة p ، التي يجب تقديرها للوحدات التي تملك الخاصّة A في المجتمع الإحصائي ، كمتوسّط متغيّرة برنولي تأخذ القيمة 1 بالنسبة للوحدات A والقيمة صفر بالنسبة للوحدات الأخرى (أنظر ص 250) . وانحراف هذه المتغيّرة النموذجي هو :

$$\sigma = \sqrt{p(1-p)} .$$

عندما يكون مقدار العيّنة كبيراً بما يكفي لجعل التقريب من القانون الطبيعي ممكناً ، نجد أنفسنا في الحالة السابقة .

لدينا ، بالنسبة للدرجة احتمال $1-\alpha$:

$$|f-p| \leqslant t_\alpha \sqrt{\frac{p(1-p)}{n}}$$

عندما يكون بوسعنا تشبيه سحب العيّنة بسحب دون ردّ .

كي تكون دقّة التقدير تساوي على الأقلّ k% من p (دقّة محدّدة بالقيمة غير المطلقة أي النسبية) ، يجب اختيار n بشكل :

$$t_\alpha \sqrt{\frac{p(1-p)}{n}} \leqslant kp$$

أي

$$n \geqslant \frac{t_\alpha^2}{k^2}\frac{1-p}{p} .$$

عند درجة احتمال ودقّة معيّنتين ، يتوقّف حجم العيّنة ، هنا ، أيضاً ، على قيمة

المتغيّر الوسيطي الذي نبحث عن تقديره . ويكون هذا الحجم أكبر كلّما كانت قيمة p أصغر أي أنّه ، كما نتوقّع ، كلّما كان عدد الوحدات A أقلّ في المجتمع الإحصائي .

يعطينا الجدول التالي حجم العيّنة n الذي يناسب ، حسب عدّة قيم لِـ p ، الدقّة k = 10% ودرجة الاحتمال $1-\alpha = 0{,}95$.

$$n = \frac{t_\alpha^2}{k^2}\frac{1-p}{p} = 400\,\frac{1-p}{p}$$

p	n
0,9	45
0,8	100
0,7	172
0,5	400
0,3	934
0,2	1 600
0,1	3 600
0,01	39 600

عملياً ، يكفي أن تكون لدينا فكرة عن مدى النسبة التي نبحث عن تقديرها كي يمكننا تحديد مقدار العيّنة بشكل معقول .

القسم II

مسائل المقارنة

1 . مبادىء اختبار الفرضيات .ـ 2 . المقارنة مع معيار : A . الاختبار المتعلّق بالتردّد ؛ B . الاختبار المتعلّق بالمتوسّط .ـ 3 . مقارنة العيّنات : A . المقارنة بين تردّدين ؛ B . المقارنة بين متوسّطين .

في كثير من الأحيان نضطر إلى مواجهة تقدير حصلنا عليه انطلاقاً من بحث إحصائي عشوائي مع معيار محدّد مسبقاً ، أو أيضاً إلى مقارنة نتائج عيّنتين مختلفتين فيما بينها . في شأن فحص المصنوعات ، نبحث مثلاً عن تحديد ما إذا كان متوسّط القطر المحسوب على عيّنة من القطع الميكانيكية المصنوعة بالجملة موافقاً للمعيار المحدّد أو ، بالعكس ، إذا كان الانحراف الملحوظ يدلّ على خلل في الآلة . خلال فحص بواسطة البحث الإحصائي لمحاسبة شركة معيّنة ، نرغب في معرفة ما إذا كان عدد الأخطاء

المبيّنة على العيّنة قابلاً للتوفيق مع النسبة المئوية للأخطاء التي تُعتبر نسبة مقبولة أم أنّ ارتفاعه بليغ . وفي دراسة حول فعالية حملة دعائية معيّنة قد نرغب ، بعد النظر إلى النتائج المسجّلة على العيّنتين A وB ، في تبيان ما إذا كانت الطريقة A أفضل ، أو لا ، من الطريقة B .

إنّ حلّ مسائل المقارنة هذه انطلاقاً من عيّنات عشوائية يستند إلى نمط تفكير إحصائي يطلق عليه إسم « اختبار الفرضيات » .

وقد التقينا بهذا النمط خلال مقارنتنا لتوزيع ملحوظ مع قانون نظري مسوّى معه (اختبار χ^2 ، الفصل III ، ص 153) .

1 . مبادىء اختبار الفرضيات

. مهما كانت المسألة المطروحة ، مراحل التفكير هي نفسها . لنضع أنفسنا ، مثلاً في حالة فحص المحاسبة بواسطة البحث الإحصائي .

لإجراء هذا التحقّق نسحب عيّنة من n مستنداً حسابياً ونعتبر نسبة p_0 من الأخطاء مقبولة . في الواقع ، إذا أردنا التأكّد مطلقاً من عدم وجود أي خطأ ، يجب القيام بفحص مستنفِد .

بشكل عام ، تكون نسبة الأخطاء الملحوظة على العيّنة مختلفة عن p_0 ، وقد تكون ، بصورة خاصّة ، أكبر منها . يمكن أن يكون سببان لهذا الانحراف :

- نسبة الأخطاء p في المحاسبة ككلّ تساوي فعلاً (أو أصغر من) p_0 والفارق الملحوظ يعود إلى مجرّد التقلّبات العشوائية ، أي الى كوننا أجرينا القياس على عيّنة ؛
- نسبة الأخطاء في المحاسبة ككلّ هي بالفعل أكبر من p_0 .

المسألة هي إذن أن نختار بين هاتين الفرضيتين ونقرّر ما إذا كان الانحراف الملحوظ معنوياً (عند درجة احتمال α محدّدة) ويدلّ على فارق حقيقي أم أنّه ، على العكس ، ليس معنوياً ويعود فقط للصدفة .

1 . نحدد الفرضيتين التبادليتين H_0 وH_1 اللتين ننوي اختبارهما :

- H_0 : نسبة الأخطاء المئوية التي تظهر في المحاسبة ككلّ تساوي النسبة المئوية المعتبرة مقبولة :

$$H_0: p = p_0$$

- H_1 : النسبة المئوية للأخطاء هي أعلى من النسبة المئوية المقبولة :

$$H_1: p > p_0$$

كان يمكننا أن نعرض فرضية أخرى H_1 : نسبة الأخطاء المئوية هي مختلفة عن النسبة المئوية المقبولة ،

$$H_1: p \neq p_0$$

ولكن في هذه الحالة يصبح طرح المسألة غير مناسب لأنّ نسبة مئوية من الأخطاء أقلّ من النسبة المئوية المقبولة تشكّل وضعاً ملائماً .

يهدف الاختبار إلى تقديم قاعدة قرار تسمح باختيار واحدة من الفرضيتين H_0 وH_1 .

2 . نعتبر الفرضية H_0 صحيحة . ضمن هذه الشروط يتحدّد قانون توزيع نسبة الأخطاء f مقاسة على العيّنة : إنّه ، حسب طريقة سحب العيّنة ، قانون ذو حدّين أو قانون فوق هندسي متوسّطها P_0 . ولا يمكن إرجاع الانحراف $f-P_0$ الملحوظ ، تحت هذه الفرضية ، إلّا إلى مجرّد تقلبات المعاينة ، أي إلى كوننا لم نجرِ الفحص إلّا على جزء من المستندات الحسابية ، وليس على مجمل المحاسبة ممّا يسبّب ، بالتالي ، بعضاً من عدم الدقّة .

3 . نحدّد درجة احتمال α ، نسميها أحياناً درجة المعنوية ، وهي عبارة عن المخاطرة التي نقبل بتحمّلها في أن نخطىء ؛ بشكل أدقّ α هي احتمال أن نأخذ H_1 فيما تكون H_0 صحيحة :(اختيار H_1 / H_0 صحيحة) $\alpha = P$

إذا أخذنا مثلاً $\alpha = 0{,}05$ فهذا يعني أنّنا نقبل 5 فرص على 100 برفض اعتبار أنّ للمحاسبة نسبة مئوية من الأخطاء أكبر من p_0 حينما تكون هذه النسبة ، في الحقيقة ، تساوي p_0 على الأكثر .

ونسب لدرجة المعنوية هذه منطقة ناقدة R احتمالها α ، ومنطقة قبول (متمّمة) $\overline{R}$ احتمالها $1-\alpha$.

4 . تنتمي نسبة الأخطاء f الملحوظة على العيّنة إمّا إلى المنطقة الناقدة R ، إمّا إلى منطقة القبول $\overline{R}$.

ويتمّ الاستدلال على الطريقة الآتية :

- f تنتمي إلى المنطقة الناقدة .

تحت الفرضية أن H_0 صحيحة ، لا يوجد سوى احتمال ضئيل α لأن نشاهد

نتيجة كهذه . إذن من المحتمل أكثر أن تكون H_0 خطئة وأن لا يكون الانحراف $f-p_0$ الملحوظ عائداً إلى مجرّد تقلّبات المعاينة فقط . بالتالي ، نرفض الفرضية H_0 ونأخذ الفرضية H_1 .

ـ f تنتمي إلى منطقة القبول .

تحت الفرضية أن H_0 صحيحة ، احتمال أن نشاهد نتيجة كهذه هو مرتفع ويساوي $1-\alpha$. إذن لا شيء يمنع من أن نقبل الفرضية H_0 . إلاّ أن هذا لا يثبت أن الفرضية الموضوعة صحيحة ، بل يعني فقط أنّ المعطيات التي بحوزتنا لا تعارض هذه الفرضية .

تُقدّم قاعدة القرار إذن على النحو التالي :

ـ إذا كانت النسبة المئوية f الملحوظة على العيّنة تنتمي إلى المنطقة الناقدة R ، نرفض الفرضية H_0 ونختار H_1 :

$f \in R$ يعني اختيار القرار H_1

ـ إذا كانت النسبة المئوية f الملحوظة على العيّنة تنتمي إلى منطقة القبول $\bar{R}$ ، نقبل الفرضية H_0 :

$f \in \bar{R}$ يعني اختيار القرار H_0.

2 . المقارنة مع معيار (Standard)

إنّ مسألة مقارنة كمّية معيّنة ، مقدّرة انطلاقاً من عيّنة ، مع قيمة محدّدة مسبقاً (معيار ، حدّ ، تخصيص ، الخ . .) هي مسألة تتردّد غالباً . ونصادفها بصورة خاصّة في إجراءات الفحص على العيّنة : النسبة المئوية للأخطاء أو الفضلات هل هي أكبر من الحدّ المفترض ، القيمة المتوسّطة لمتغيّر وسيطي معيّن (قطر قطعة ميكانيكية ، مدة حياة عنصر الكتروني ، الخ . .) هل تساوي القيمة المحدّدة ؟

إنّ هذه المسألة ، مقارنة قيمة مقياس θ مع معيار θ_0 ، تستدعي اختبار فرضيتين تبادليتين H_0 وH_1 . وبإمكان الفرضية H_1 أن تأخذ أشكالاً مختلفة حسب طبيعة المسألة المطروحة :

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 . \end{cases} \qquad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 \end{cases} \qquad \begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

يؤدّي كلّ من هذه الحالات الثلاث إلى قواعد اختبار مختلفة : في الحالة الأولى ، تكون المنطقة الناقدة بأكملها إلى يمين فسحة تغيّر θ [1] ؛ في الحالة الثانية ، تكون بأكملها إلى اليسار ؛ وفي الثالثة موزّعة بالتماثل على يمين ويسار فسحة التغيّر .

(1) نتّبع اتجاه الكتابة اللاتينية .

A . الاختبار المتعلّق بالتردّد

لنأخذ مجتمعاً إحصائياً مؤلّفاً من وحدات يمتلك قسم منها الخاصّة A . سحبنا من هذا المجتمع عيّنة حجمها n ولاحظنا عليها التردّد f بالنسبة للوحدات التي لها هذه الخاصّة .

النسبة p للوحدات A في المجتمع الإحصائي هي مجهولة وقد تختلف f عنها بحكم تقلّبات المعاينة . على أساس القيمة الملحوظة f ننوي اختبار ما إذا كان يمكن اعتبار p ، أو لا يمكن ، مساوية لقيمة p_0 محدّدة مسبقاً .

1 ـ نحدّد تبعاً للمسألة المطروحة الفرضيتين التبادليتين H_0 وH_1 اللتين نرغب في اختبارهما ، ونجد أنفسنا في واحدة من الحالات الثلاث :

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 \end{cases} \qquad \begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 \end{cases} \qquad \begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 . \end{cases}$$

2 . يتبع التردّد f ، حسب طريقة سحب العيّنة ، قانوناً ذا حدين أو قانوناً فوق هندسي متغيّره الوسيطي ، إذا اعتبرنا الفرضية H_0 صحيحة ، $p = p_0$.

ضمن عدد من الشروط ، غالباً ما تتحقّق عملياً ـ مقدار العيّنة n كبير بشكل كاف ، أو ، بالنسبة لعيّنة مستنفدة ، نسبة البحث الاحصائي n/N ضعيفة ـ يحقّ لنا تقريب هذين القانونين من قانون طبيعي متوسّطه $p=p_0$ وانحرافه النموذجي

$\sigma_0 = \sqrt{p_0(1 - p_0)/n}$ (1).

إذن المتغيّرة

$$T = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

تتبع قانوناً طبيعياً ممركزاً مختصراً .

---

(1) إذا لم تتحقّق هذه الشروط ، يجب استعمال القانون الصحيح : القانون ذا الحدّين ، القانون فوق الهندسي ، قانون بواسّون أو أيضاً التقريب من القانون الطبيعي ذي الانحراف النموذجي

$\sigma_0 = \sqrt{p_0(1 - p_0)/n} . \sqrt{(N - n)/(N - 1)}$

حسب الحالة (انظر ، تقدير النسبة ، القسم I ، الفقرة 2.B) .

3 . عندما نعرف درجة المعنوية α ، يمكننا تحديد المنطقة الناقدة التي تناسب كلاً من الحالات الثلاث السابقة .

الحالة الأولى

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p > p_0 . \end{cases}$$

المنطقة الناقدة هي بالشكل : f > l ، ونحدّد قيمة l بشكل يكون فيه ·

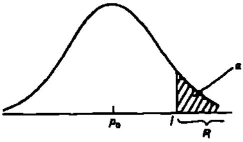
P { اختيار H_1 / H_0 صحيحة } $= P\{f > l / p = p_0\} = \alpha$ (أنظر الشكل 55)

ونقرأ في الجدول $\pi(t)$ أو $p(t)$ قيمة المتغيّرة الطبيعية المركزة المختصرة t_α حيث

$$P\{T > t_\alpha\} = \alpha .$$

ونستنتج قيمة l :

$$l = p_0 + t_\alpha \sqrt{\frac{p_0(1 - p_0)}{n}} .$$



الشكل 55

قاعدة الاختبار هي التالية :

إذا كان التردّد الملحوظ f أكبر من l ، نرفض الفرضية H_0 لأنّ احتمال قيمة مرتفعة بهذا الشكل لـ f، تحت الفرضية H_0 ، هو احتمال ضعيف :

$f > l$ يعني اختيار القرار H_1

في الحالة المعاكسة نقبل الفرضية H_0 :

$f < l$ يعني اختيار القرار H_0

الحالة الثانية :

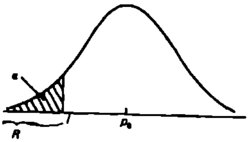
$$\begin{cases} H_0 : p = p_0 \\ H_1 : p < p_0 . \end{cases}$$

المنطقة الناقدة R هي بالشكل f<1 ، ونحدّد قيمة 1 ، بشكل يكون فيه :

P { اختيار H_1 / H_0 صحيحة } $= P\{f < 1 / p = p_0\} = \alpha$ (أنظر الشكل 56) ومن

قيمة ما حيث :

$P\{T < t_\alpha\} = \alpha$



الشكل 56

والتي نقرأها في الجدول ، نستنتج كما في السابق قيمة l

نصل إلى قاعدة الاختبار :

إذا كان التردّد الملحوظ f أصغر من l ، نرفض الفرضية H_0 :

$f < l$ يعني اختيار القرار H_1 .

ونقبل H_0 في الحالة المعاكسة :

$f > l$ يعني اختيار القرار H_0 .

الحالة الثالثة :

$$H_0 : p = p_0$$
$$H_1 : p \neq p_0 .$$

هذه المرّة ، منطقة القبول $\bar{R}$ هي منطقة متماثلة (متناظرة) شكلها :

$$l_1 < f < l_2$$

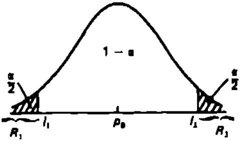
ونحدّد القيمتين l_1 وl_2 بشكل يكون فيه :

$P\{$ اختيار H_0/H_0 صحيحة $\} = P\{l_1 < f < l_2/p = p_0\} = 1 - \alpha$ (أنظر الشكل 57) .

وتتكوّن المنطقة الناقدة R من قسمين متماثلين R_1 وR_2 يوافق كلاً منها الاحتمال $\alpha/2$.

نقرأ في الجدول $\Pi(t)$ أو $P(t)$ قيمة المتغيّرة الطبيعية المركزة المختصرة $t_{\alpha/2}$ حيث :

$$P\{T > t_{\alpha/2}\} = \frac{\alpha}{2}.$$



الشكل 57

ونستنتج قيمة حدّي منطقة القبول l_1 وl_2 :

$$l_1 = p_0 - t_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}, \qquad l_2 = p_0 + t_{\alpha/2}\sqrt{\frac{p_0(1-p_0)}{n}}.$$

إذن قاعدة الاختبار هي التالية :

إذا كان التردّد الملحوظ f خارج الفسحة (l_1, l_2) ، نرفض الفرضية H_0 . :

$$\left\{ \begin{array}{l} f < l_1 \\ f > l_2 \end{array} \right.$$ يعني اختيار القرار H_1 .

ونقبل H_0 في الحالة المعاكسة :

$l_1 < f < l_2$ يعني اختيار القرار H_0 .

مثلاً : ننوي بواسطة البحث الإحصائي أن نفحص دقّة عملية جرد بضاعة تجارية تتضمّن عشرة آلاف سلعة. نسحب عيّنة من 500 سلعة لهذا الهدف ونعتبر أن نسبة الأخطاء في عملية الجرد مقبولة إذا كانت أصغر من أو تساوي 3% .

في هذا المثل

N كبيرة جدّاً ، $n = 500$ و$p_0 = 0,03$

الفرضيتان التبادليتان اللتان ننوي اختبارهما هما :

$$H_0 : p = 0,03 \qquad , \qquad H_1 : p > 0,03$$

يتبع التردّد f الملحوظ على العيّنة قانوناً ذا حدّين إذا تمّ سب العيّنة مع ردّ ، أو قانوناً فوق هندسي في الحالة ، التي غالباً ما تتكرّر عملياً ، حيث يكون سحب العيّنة دون ردّ . وفي كلتي الحالتين ، يمكننا تقريب هذين القانونين بقانون طبيعي (معتدل) . إذا افترضنا H_0 صحيحة ، فمتوسّط هذا القانون هو $p = p_0$ وانحرافه النموذجي $\sigma_0 = \sqrt{p_0(1 - p_0)/n}$.

ويصبح شكل المنطقة الناقدة : ب $f > l$ حيث :

$$P\{ f > l / p = p_0 \} = \alpha .$$

إذا أخذنا درجة المعنوية $\alpha = 0,05$ ، فقيمة المتغيّرة الطبيعية المركزة المختصرة t_α التي نقرؤها في الجدول حيث :

$$P\{ T > t_\alpha \} = \alpha ,$$

هي :

$$t_{0,05} = 1,65 .$$

بالتالي :

$$l = p_0 + t_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} = 0{,}03 + 1{,}65 \sqrt{\frac{0{,}03 \times 0{,}97}{500}}$$
$$= 0{,}03 + 1{,}65 \times 0{,}007\,6 = 0{,}043\,.$$

إذن نرفض الفرضية ونعتبر أنّ نسبة الأخطاء المرتكبة في عملية الجرد أكبر من 3% معنوياً إذا كانت نسبة الأخطاء المئوية المأخوذة على العيّنة أكبر من 4,3% .

B . الاختبار المتعلّق بالمتوسّط

لاحظنا على عيّنة حجمها n ، القيمة المتوسّطة $\bar{x}$ بالنسبة لمتغيّرة إحصائية X.

قيمة المتوسّط m الحقيقية بالنسبة لمجمل المجتمع الإحصائي هي مجهولة وقد تختلف $\bar{x}$ عنها بحكم التقلّبات العشوائية . على أساس القيمة الملحوظة $\bar{x}$ ، ننوي اختبار ما إذا كان يمكن اعتبار المتوسّط m ، أو لا يمكن ، مساوياً لقيمة m_0 محدّدة مسبقاً .

نمط التفكير هو نفسه كما بالنسبة للتردّد ، والصعوبة الوحيدة تكمن في كون الانحراف النموذجي σ للمتغيّرة الإحصائية X غير معروف بشكل عام إلاّ من خلال القيمة التي نجدها على العيّنة .

1 . تبعاً للمسألة المطروحة ، نحدّد الفرضيتين التبادليتين H_0 وH_1 اللتين قد تكونان ، حسب الحالة :

$$\begin{cases} H_0 : m = m_0 \\ H_1 : m > m_0 \end{cases} \qquad \begin{cases} H_0 : m = m_0 \\ H_1 : m < m_0 \end{cases} \qquad \begin{cases} H_0 : m = m_0 \\ H_1 : m \neq m_0\,. \end{cases}$$

2 . إذا كان المجتمع الإحصائي الأصل هو نفسه موزّعاً حسب القانون الطبيعي أو إذا كان مقدار العيّنة كبيراً بدرجة كافية ، أكثر من ثلاثين وحدة ، فإنّ $\bar{x}$ تتبع تماماً أو تقريباً قانون لابلاس - غوس (Laplace-Gauss) بمتغيرين وسيطين m و$\sigma/\sqrt{n}$ [1] ، حيث m وσ هما متوسّط المتغيّرة الإحصائية X وانحرافها النموذجي في مجمل المجتمع الإحصائي .

---

(1) أو $\frac{\sigma}{\sqrt{n}} \cdot \sqrt{(N-n)/(N-1)}$ في حالة سحب مستنفد لا يمكننا تشبيهه بسحب مستقل بحكم القيمة المرتفعة لنسبة البحث الإحصائي n/N .

إنّ اختبار الفرضية H_0 $(m=m_0)$ صحيحة لا يكفي إذن لتحديد قانون احتمال $\bar{x}$ كليّاً : فهذا القانون يتعلّق بقيمة σ التي قد تكون ، حسب الحالة ، معروفة أو غير معروفة .

3 . الانحراف النموذجي σ معروف . قليلاً ما نلتقي بهذه الحالة عملياً ، المتغيّرة :

$$T = \frac{\bar{x} - m_0}{\sigma/\sqrt{n}}$$

تتبع قانوناً طبيعياً ممركزاً مختصراً .

يمكننا عندما نحدّد درجة المعنوية α أن نعيّن المنطقة الناقدة التي تناسب كلاًّ من الحالات الثلاث السابقة .

مثلاً خلال اختبار الفرضية : $H_0 : m = m_0$

مقابل $H_1 : m \neq m_0$

تكون منطقة القبول بالشكل :

$$l_1 < \bar{x} < l_2$$

حيث نحدّد القيمتين l_1 وl_2 بشكل يكون فيه :

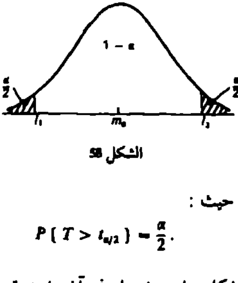
$$P \{ \text{اختيار } H_0 \backslash H_0 \text{ صحيحة} \}$$

$$= P \{ l_1 < \bar{x} < l_2 / m = m_0 \} = 1 - \alpha$$

(أنظر الشكل 58) .

نحدّد إذن منطقة القبول بواسطة :

$$m_0 - t_{\alpha/2} . \frac{\sigma}{\sqrt{n}} < \bar{x} < m_0 + t_{\alpha/2} . \frac{\sigma}{\sqrt{n}}$$



الشكل 58

$t_{\alpha/2}$ هي قيمة المتغيّرة الطبيعية الممركزة المختصرة حيث :

$$P \{ T > t_{\alpha/2} \} = \frac{\alpha}{2} .$$

4 ـ الانحراف النموذجي σ مجهول . بشكل عام ، نجهل في آن واحد قيمة المتوسّط المجتمع الإحصائي وانحرافه النموذجي . عندئذٍ نعتمد مكان σ تقديرها s' الذي نستنتجه من المشاهدات :

$$s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 .$$

إذا كان مقدار العيّنة كبيراً ، أكثر من 30 يكون s'^2 تقديراً لِـ σ^2 دقيقاً بشكل كاف كي تكون المتغيّرة المركزة المختصرة التي استبدلنا في حسابها σ بواسطة s'

$$T = \frac{\bar{x} - m_0}{s'/\sqrt{n}}$$

موزّعة حسب القانون الطبيعي . وهكذا نعود إلى الحالة حيث الانحراف النموذجي معروف .

بالمقابل ، إذا كان المقدار n صغيراً ، أقلّ من 30 ، لا يمكننا ، بحكم تقلّبات مخرج T العشوائية أن نشبّهها بمتغيّرة طبيعية ممركزة مختصرة. إنّها تتبع قانون ستودنت ـ فيشر (Student-Fisher) بِـ $n-1$ درجة حرّية . وهكذا نضطرّ ، لتحديد منطقة القبول ، إلى استعمال قانون ستودنت بدلاً من قانون لابلاس ـ غوس .

مثل 1 . تصنع إحدى الآلات قطعاً ميكانيكية بالجملة ، وقد ضُبطت كي يكون قطر هذه القطع يساوي 12,60 ملم . طبعاً لا بدّ من بعض قابلية للتغيّر . لاحظنا على عيّنة من 100 قطعة قيمة متوسّطة لهذا القطر $\bar{x}$ تبلغ $\bar{x} = 12{,}65$ mm وتبايناً $s^2=0{,}1584$. هل يمكن اعتبار ضبط الآلة صحيحاً ؟

في هذا المثل ، ننوي اختبار الفرضية $H_0: m = 12{,}60$.

مقابل $H_1 : m \neq 12.60$.

إنّ حجم العيّنة كبير كاف لجعل المتوسّط الملحوظ يتبع قانوناً طبيعياً متوسّطه m وانحرافه النموذجي $\sigma/\sqrt{n}$. قيمة σ الحقيقية مجهولة ولكن يحقّ لنا تقديرها بواسطة s':

$$s'^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{n}{n-1}s^2 = \frac{100}{99} \times 0.1584 = 0.1600$$

$$s' = 0.40 .$$

إذا اعتبرنا الفرضية H_0 صحيحة ، فإنّ المتغيّرة :

$$T = \frac{\bar{x} - m_0}{s'/\sqrt{n}} = \frac{\bar{x} - 12.60}{0.04}$$

هي موزّعة حسب القانون الطبيعي .

بحكم الفرضيتين التبادليتين المأخوذتين ، منطقة القبول هي على الشكل :

$$l_1 < \bar{x} < l_2$$

حيث :

$$P\{ l_1 < \bar{x} < l_2 / m = m_0 \} = 1 - \alpha .$$

إذا أخذنا درجة المعنوية α=0,05 ، فإنّ قيمة المتغيّرة الطبيعية المركزة المختصرة $t_{\alpha/2}$ التي نقرؤها في الجدول حيث

$$P\{ T > t_{\alpha/2} \} = \alpha/2$$

هي

$$t_{0.025} = 1,96 \simeq 2 .$$

بالتالي :

$$l_1 = m_0 - t_{\alpha/2} \frac{s'}{\sqrt{n}} = 12,60 - 2 \times 0,04 = 12,52$$

$$l_2 = m_0 + t_{\alpha/2} \frac{s'}{\sqrt{n}} = 12,60 + 2 \times 0,04 = 12,68 .$$

إذن منطقة القبول هي :

$$12,52 < \bar{x} < 12,68$$

القيمة الملحوظة ($\bar{x}$ = 12,65) توجد ضمن هذه المنطقة ، إذن هي لا تعارض الفرضية H_0 : لا تسمح لنا القياسات التي أجريناها على العيّنة بوضع صحّة ضبط الآلة موضع الشك .

مثل 2 . لنفترض أنّه في المثل السابق لاحظنا القيمة المتوسّطة $\bar{x}$ = 12,65 mm والتباين s^2=0,1584 على عيّنة من 10 قطع فقط .

ضمن هذه الشروط :

$$s'^2 = \frac{n}{n-1} s^2 = \frac{10}{9} \times 0,158\,4 = 0,176$$

$$\frac{s'}{\sqrt{n}} = \sqrt{\frac{0,176}{10}} = 0,13 .$$

بما أنّ حجم العيّـنة ضعيف ، s' هو تقدير غير كاف للانحراف النموذجي σ كي يمكن اعتبار المتغيّـرة :

$$T = \frac{\bar{x} - m_0}{s'/\sqrt{n}} = \frac{\bar{x} - 12{,}60}{0{,}13}$$

موزّعة طبيعياً . إنّـها تتبع قانون ستودنت ـ فيشر $n-1=9$ درجات حرّية .

بالنسبة للدرجة المعنوية $\alpha=0{,}05$ ، القيمـة $t_{\alpha/2}$ التي نقرؤها في جدول ستودنت فيشر (الملحق : الجدول 6) لِـ 9 درجات حرّية ، حيث

$$P\{ T > t_{\alpha/2} \} = \frac{\alpha}{2}$$

هي

$$t_{0.025} = 2{,}26$$

ومنطقة القبول هي :

$$m_0 - t_{\alpha/2}\frac{s'}{\sqrt{n}} < \bar{x} < m_0 + t_{\alpha/2}\frac{s'}{\sqrt{n}}$$
$$12{,}60 - 2{,}26 \times 0{,}13 < \bar{x} < 12{,}60 + 2{,}26 \times 0{,}13$$
$$12{,}31 < \bar{x} < 12{,}89 .$$

منطقة القبول الجديدة هي إذن أوسع من سـابقتها : في الـواقع ، بمـا أنّ مقدار العيّـنة أضعف ، يمكن لمجرّد تقلّبات المعاينة أن تفسّـر انحرافات أكبر دون أن نحتاج للشـكّ بصحـة الفـرضيـة H_0 . يضـاف إلى هـذا شـكّ متزايـد في تقييم الانحـراف النموذجي .

3 . مقارنة العيّـنات

يتجه عدد كبير من المسائل التقنية أو التجارية ، كتحليل أخير ، إلى مقارنة بـين النتائج الحاصلة على عيّـنات مختلفة . بين نهجي صناعة ، أيّـهما يعطي نسـبة فضـلات أقل ؟ هل تتيح الوسيلة الدعائية A بالوصول إلى عدد من الأفراد أكثر أو أقلّ ارتفاعاً من الوسيلة B ؟ هل زاد متوسّط استهلاك منتوج معيّـن أو تناقص بـين الفترة 1 والفتـرة 2 غالباً ما يتمّ ، في الواقع ، حلّ هذا النوع من المسـائل عـلى أساس دراسـات بواسـطة البحث الإحصائي .

لنأخذ مجتمعين إحصائيين P_1 وP_2 نأخذ منهما عيّـنتين قد يكون حجماهما مختلفين .

ننوي انطلاقاً من النتائج الملحوظة على العيّنتين أن نقرّر ما إذا كان يمكن اعتبار قيمتي مقياس معيّن θ متساويتين أو مختلفتين في المجتمعين .

عادة تكون القيمتان مختلفتين ، ويمكن نسب هذا الاختلاف إلى سببين :

- القيمتان θ_1 وθ_2 هما بالفعل مختلفتان في المجتمعين الإحصائيين ،
- قيمتا المقياس θ_1 وθ_2 موضع الدراسة هما نفسهما في المجتمعين الاحصائيين والفارق الملحوظ يعود إلى مجرّد تقلّبات المعاينة .

علينا الاختيار بين هاتين الفرضيتين . تؤدّي المسألة إلى اختبار الفرضية :

$$H_0 : \theta_1 - \theta_2 = 0$$

التي نطلق عليها عامّة اسم الفرضية الصفر ، مقابل الفرضية البديلة :

$$H_1 : \theta_1 - \theta_2 \neq 0 .$$

علينا إذن أن نشكّل الفارق بين النتائج الملحوظة على العيّنتين وأن نتساءل إن كان هذا الفارق معنوياً (كاشفاً) أم لا .

خصائص الفارق بين متغيّرتين عشوائيتين

لنتذكّر بعض الخصائص المتعلّقة بالفارق بين متغيّرتين عشوائيتين .

لنفترض X_1 وX_2 متغيّرتين عشوائيتين مستقلّتين ولنأخذ الفارق بينهما $X_1 - X_2$.

1 . أمل فارق المتغيّرتين العشوائيتين الرياضي يساوي الفارق بين الأملين الرياضيين (أنظر الفصل I ، ص 57) .

$$E\{X_1 - X_2\} = E\{X_1\} - E\{X_2\} .$$

2 . تباين فارق متغيّرتين عشوائيتين مستقلّتين يساوي مجموع التباينين (أنظر الفصل I ، ص 61) .

$$V\{X_1 - X_2\} = V\{X_1\} + V\{X_2\} .$$

بالتالي :

$$\sigma_{X_1 - X_2} = \sqrt{\sigma^2_{X_1} + \sigma^2_{X_2}} .$$

3 إذا كانت المتغيّرتان X_1 وX_2 موزّعتين حسب قانونين طبيعيين متغيراتها الوسيطية على التوالي :

$m_1, \sigma_1 \quad , \quad m_2, \sigma_2$

يكون الفارق $(X_1 - X_2)$ نفسه موزّعاً حسب قانون طبيعي بمتغيّرين وسيطين :

$$E\{X_1 - X_2\} = m_1 - m_2$$
$$\sigma_{X_1 - X_2} = \sqrt{\sigma_1^2 + \sigma_2^2}.$$

A . المقارنة بين تردّدين

لنأخذ مجتمعين إحصائيين P_1 وP_2 يتألّفان من وحدات يمتلك بعضها الخاصّة A في كلّ من المجتمعين p_1 وp_2 هما مجهولتان .

نأخذ :

- عيّنة حجمها n_1 من p_1 ،
- عيّنة حجمها n_2 من p_2 .

على هاتين العيّتين نلاحظ على التوالي التردّدين f_1 وf_2 بالنسبة للوحدات A . ننوي على أساس هذه المشاهدات أن نقرّر ما إذا كان يمكن اعتبار النسبتين p_1 وp_2 الموجودتين في المجتمعين ، متساويتين.

1 . الفرضيتان التبادليتان اللتان نرغب في اختبارهما هما :

$$\begin{cases} H_0 : p_1 - p_2 = 0 \\ H_1 : p_1 - p_2 \neq 0. \end{cases}$$ (الفرضية الصفر)

2 . يتبع التردّدان ، حسب طريقة سحب العيّتين ، قوانين ذات حدّين أو فوق هندسية . إذا كان المقداران n_1 وn_2 كبيرين بدرجة كافية يصحّ التقريب من القانون الطبيعي . في هذه الظروف وبشرط أن يكون بالإمكان تشبيه سحبي العيّنة بسحبين مستقلّين[1] :

- يتبع التردّد f_1 قانوناً طبيعياً متغيّراه الوسيطيان :

---

(1) إن لم يكن الحال كذلك ، يصبح الانحراف النموذجي للقانون الطبيعي الذي تتبعه f_1 على الشكل :

$$\sigma_1 = \sqrt{p_1(1 - p_1)/n_1} \cdot \sqrt{(N_1 - n_1)/(N_1 - 1)}$$

كذلك بالنسبة للتردّد f_2 .

$$m_1 = p_1 , \qquad \sigma_1 = \sqrt{\frac{p_1(1-p_1)}{n_1}};$$

- ويتبع التردّد f_2 قانوناً طبيعياً متغيّراه الوسيطيان :

$$m_2 = p_2 . \qquad \sigma_2 = \sqrt{\frac{p_2(1-p_2)}{n_2}} .$$

بناء على الخاصتين المذكورتين أعلاه ، يتبع الفارق $d=f_1-f_2$ قانوناً طبيعياً متغيّراه الوسيطيان :

$$m = E\{d\} = E\{f_1\} - E\{f_2\} = p_1 - p_2$$
$$\sigma = \sigma_d = \sqrt{\sigma_{f_1}^2 + \sigma_{f_2}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} .$$

لنعتبر أنّ الفرضية H_0 :

$$H_0 : p_1 - p_2 = 0 \qquad , \qquad p_1 = p_2 = p$$

هي صحيحة . تحت هذه الفرضية يتبع الفارق d قانوناً طبيعياً :

$$\mathcal{N}\left\{0, \sqrt{p(1-p)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}\right\} ,$$

حيث p تمثّل قيمة p_1 وp_2 المشتركة .

3 . إذا كنّا نعرف درّجة المعنوية α ، يمكننا تحديد فسحة القبول المتماثلة :

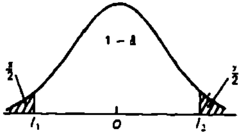
$$l_1 < d < l_2$$

المعيّنة بواسطة :

$P\{H_0 \backslash H_0 \text{ اختيار } \text{ صحيحة}\} = P\{l_1 < d < l_2 / p_1 = p_2\} = 1 - \alpha$ (أنظر الشكل 59) .

ونحصل على :

$$- t_{\alpha/2}\,\sigma_d < d < + t_{\alpha/2}\,\sigma_d$$



الشكل 59

$t_{\alpha/2}$ هي قيمة المتغيّرة الطبيعية المركزة المختصرة

حيث :
$$P\{T > t_{\alpha/2}\} = \alpha/2 .$$

في الحقيقة ، σ_d هي قيمة مجهولة :

$$\sigma_d = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} ;$$

نقدِّر p ، القيمة المشتركة لِـ p_1 وp_2 ضمن الفرضية H_0 ، بواسطة التردّد f المحسوب على مجموع العيّنتين . إذا أشرنا بواسطة x_1 وx_2 إلى عدد الوحدات A الملحوظة على كلّ من العيّنتين :

$$f = \frac{x_1 + x_2}{n_1 + n_2} = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} ;$$

إذاً ، نقدّر σ_d بواسطة :

$$s_d = \sqrt{f(1-f)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

ونحدّد أخيراً فسحة القبول ، عند درجة المعنوية α ، بواسطة :

$$- t_{\alpha/2}\, s_d < d < + t_{\alpha/2}\, s_d .$$

يمكننا التعبير عن هذه الفسحة تبعاً لخارج القسمة d/s_d :

$$- t_{\alpha/2} < d/s_d < + t_{\alpha/2} .$$

مثلاً . لتحديد نسبة شَغْل عتاد باهظ ، نعتمد طريقة « المشاهدات الآنية » : على طول كلّ شهر نلاحظ عيّنة لحظات مسحوبة بالصدفة . عند كلّ من هذه اللحظات المحدّدة يسجّل مراقب ما إذا شُغل العتاد أو لا . بهذه الطريقة لاحظنا عيّنة من 500 لحظة في شهر كانون الثاني (يناير) ومن 400 لحظة في شهر شباط (فبراير) . وحصلنا على النتائج الآتية :

	كانون الثاني	شباط
شغل	400	300
عدم شغل	100	100
المجموع	500	400

هل يوجد فارق معنوي (كاشف) بين شغل هذا العتاد في كانون الثاني وشباط ؟
في هذا المثل :

$$n_1 = 500\,, \qquad f_1 = \frac{400}{500} = 0{,}80$$

$$n_2 = 400\,, \qquad f_2 = \frac{300}{400} = 0{,}75\,.$$

في الفرضية الصفر :

$$H_0 : p_1 = p_2 = p\,, \qquad p_1 - p_2 = 0\,,$$

يتبع الفارق $d = f_1 - f_2$ قانوناً طبيعياً متوسّطه $m = 0$ وانحرافه النموذجي :

$$\sigma_d = \sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

يتمّ تقدير p بواسطة :

$$f = \frac{n_1 f_1 + n_2 f_2}{n_1 + n_2} = \frac{400 + 300}{900} = 0{,}78\,.$$

إذن نقدّر σ_d بواسطة :

$$s_d = \sqrt{0{,}78 \times 0{,}22\left(\frac{1}{500} + \frac{1}{400}\right)} = 0{,}028\,.$$

تتناسب درجة المعنوية $\alpha = 0{,}05$ مع القيمة

$$t_{\alpha/2} \doteq 2$$

إذن فسحة قبول الفرضية H_0 هي :

$$-2 \times 0{,}028 < d < +2 \times 0{,}028$$

$$-0{,}056 < d < +0{,}056$$

الفارق الملحوظ

$$d = f_1 - f_2 = 0{,}05$$

هو موجود ضمن هذه الفسحة : إنّه ليس معنوياً . لا تسمح لنا المشاهدات التي بحوزتنا أن نؤكّد أنّ نسبة شغل العتاد قد تضاءلت في شهر شباط : يمكننا نسب الفارق الملحوظ بين الشهرين فقط إلى مجرّد تقلّبات المعاينة .

B . المقارنة بين متوسّطين

لنأخذ مجتمعين إحصائيين P_1 وP_2 ونسحب :

- عيّنة حجمها n_1 من P_1 ،
- عيّنة حجمها n_2 من P_2 .

لنفترض $\bar{x}_1$ و$\bar{x}_2$ متوسّطي المتغيّرة الإحصائية X في كلّ عيّنة .ننوي على أساس هذه المشاهدات اختبار ما إذا كان متوسّط المتغيّرة X هو نفسه في المجتمعين أو لا .

لنرمز على التوالي بواسطة :

m_1, σ_1 ، m_2, σ_2 ،

إلى متوسّط X وانحرافها النموذجي في P_1 وP_2 .

1 . الفرضيتان التبادليتان اللتان نرغب في اختبارهما هما :

$$H_0 : m_1 - m_2 = 0 \quad \text{(الفرضية الصفر)}$$
$$H_1 : m_1 - m_2 \neq 0 .$$

2 . إذا كانت المتغيّرة الإحصائية X موزّعة في كلّ مجتمع إحصائي حسب القانون الطبيعي ، فإنّ المتوسّطين $\bar{x}_1$ و $\bar{x}_2$ يتبعان بدورهما قانوناً طبيعياً .

إلّا أنّه إذا لم يبدُ افتراض التوزيع الطبيعي في المجتمعين مبرّراً ، يكفي أن يكون مقدارا العيّنتين n_1 وn_2 كبيرين بدرجة كافية (أكثر من 30 وحدة تقريباً) كي يكون توزيعا $\bar{x}_1$ و$\bar{x}_2$ تقريباً طبيعيين .

ضمن هذه الشروط العامّة جداً وإذا افترضنا أنّه يمكن تشبيه سحبي العيّنتين بسحبين مستقلّين[1] :

- $\bar{x}_1$ يتبع قانوناً طبيعياً متغيّراه الوسيطيان :

$$E(\bar{x}_1) = m_1 , \qquad \sigma_{\bar{x}_1} = \sqrt{\frac{\sigma_1^2}{n_1}}$$

(1) إذا لم يكن الحال كذلك ، يصبح الانحراف النموذجي للقانون الذي يتبعه $\bar{x}_1$:

$$\sigma_{\bar{x}_1} = \sqrt{\sigma_1^2/n_1} . \sqrt{(N_1 - n_1)/(N_1 - 1)} .$$

كذلك بالنسبة لـ $\bar{x}_2$

- $\bar{x}_2$ يتبع قانوناً طبيعياً متغيّراه الوسيطيان :

$$E\{\bar{x}_2\} = m_2, \qquad \sigma_{\bar{x}_2} = \sqrt{\frac{\sigma_2^2}{n_2}}.$$

بالتالي ، بفضل الخصائص المذكورة أعلاه (ص 291) ، يتبع الفارق $d = \bar{x}_1 - \bar{x}_2$ هو أيضاً قانوناً طبيعياً متغيّراه الوسيطيان :

$$E\{d\} = m_1 - m_2, \qquad \sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

لنعتبر الافتراض :

$$H_0 : m_1 - m_2 = 0$$

صحيحاً . تحت هذه الفرضية ، توزيع احتمال d هو قانون طبيعي :

$$\mathcal{N}\left\{0, \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right\}.$$

إنّ اعتبار الفرضية H_0 متحقّقة لا يكفي إذن لتحديد قانون إحتمال d كلّياً : فهذا القانون يتعلّق بقيمتي σ_1 وσ_2 اللتين قد تكونان ، حسب الحالة ، معروفتين أو مجهولتين .

3 . σ_1 وσ_2 معروفتان . بإعطائنا درجة المعنوية α ، نحدّد منطقة القبول بواسطة :

$$-t_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < d < +t_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}},$$

$t_{\alpha/2}$ هي قيمة المتغيّرة الطبيعية المركزة المختصرة حيث :

$$P\{T > t_{\alpha/2}\} = \frac{\alpha}{2}.$$

4 . σ_1 وσ_2 غير معروفتين . نضطر في هذه الحالة إلى تقدير :

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

باستبدالنا σ_1^2 وσ_2^2 بواسطة تقديرهما :

$$s_1'^2 = \frac{1}{n_1 - 1}\sum_i (x_{1i} - \bar{x}_1)^2$$

$$s_2'^2 = \frac{1}{n_2 - 1}\sum_j (x_{2j} - \bar{x}_2)^2$$

إذا كان مقدارا العينتين n_1 وn_2 كبيرين بدرجة كافية ، إذن

$$s_d' = \sqrt{\frac{s_1'^2}{n_1} + \frac{s_2'^2}{n_2}}$$

هي تقدير كاف لـ σ_d . يمكننا إذن ، تحت الفرضية H_0 ، الاعتبار أنّ المتغيّرة المركزة المختصرة :

$$T = \frac{d}{s_d'}$$

تتبع تقريباً قانوناً طبيعياً ونعود إلى الحالة 3 . حيث يكون التباينان معروفين .

بالمقابل ، عندما يكون مقدارا العيّنتين ضعيفين ، لا يكون التقديران $s_1'^2$ و $s_2'^2$ دقيقين وقد يختلفان بشكل ملموس عن القيمتين الحقيقيتين σ^2_1 وσ^2_2 . ضمن هذه الشروط لا يعود تطبيق الاختبار السابق ممكناً : فهو لا يسمح بتمييز ما إذا كان يمكن نسب الفارق الملحوظ بين المتوسّطين $\bar{x}_1$ و$\bar{x}_2$ إلى اختلاف حقيقي بين المتوسّطين m_1 وm_2 أم إلى اختلاف بين التباينين σ^2_1 وσ^2_2 . لم يتمّ التوصّل إلى حلّ موافق تماماً بالنسبة لهذه المسألة[1] .

مثلاً . أجري في تجمّع سكّاني كبير ، تحقيق بواسطة البحث الإحصائي حول نفقات الأسر الشهرية على المآكل . كانت العيّنة تتضمّن 327 أسرة من العمّال و286 أسرة من الموظفين . وقد لاحظنا القيم التالية المتعلّقة بمتوسّط الاستهلاك الغذائي وانحرافه النموذجي في هاتين الفئتين الاجتماعيتين .

	المقدار	المتوسّط	الانحراف النموذجي
عمّال	$n_1 = 327$	$\bar{x}_1 = 612$ F	$s_1 = 104$ F
موظّفون	$n_2 = 286$	$\bar{x}_2 = 642$ F	$s_2 = 118$ F

(1) يمكننا حول هذا الموضوع مراجعة : G. Darmois « مقارنة متوسّطي مجتمعين إحصائيين طبيعيين بانحرافين نموذجيين مجهولين ومختلفين » . نشرة الإحصاء التطبيقي ، المجلّد 2 ، العدد 3 ، 1954

هل يمكننا الاستنتاج أن أسر الموظّفين تنفق على المأكل أكثر من أسر العمّال ؟

في دراسة من هذا النوع ، حتى ولو تمّ سحب العيّنة حتماً دون ردّ ، يمكننا تشبيهها بعيّنة مسحوبة مع ردّ (سحوبات مستقلّة) بحكم ضعف نسبة البحث الإحصائي : فالتجمّع السكّاني الكبير يحتوي على عشرات الآلاف من الأسر .

لنرمز على التوالي بواسطة :

$$m_1, \sigma_1, \qquad m_2, \sigma_2$$

إلى متوسّط الاستهلاك الغذائي وانحرافه النموذجي في مجموعة أسر العمّال وأسر الموظّفين التي تنتمي إلى التجمّع السكّاني .

في الفرضية الصفر :

$$H_0 : m_1 - m_2 = 0,$$

يتبع الفارق $d = \bar{x}_1 - \bar{x}_2$ قانوناً طبيعياً متوسّطه $E\{d\} = 0$ ، وانحرافه النموذجي :

$$\sigma_d = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\,.$$

نقدّر σ_d باستبدالنا σ^2_1 وσ^2_2 بواسطة تقديرهما انطلاقاً من العيّنة . مقدارا أسر العمّال والموظّفين n_1 وn_2 الممثّلة في العيّنة هما كبيران بشكل كاف كي يكون :

$$s_1'^2 \# s_1^2 \qquad s_2'^2 \# s_2^2 .$$

يكفي إذن استبدال σ^2_1 وσ^2_2 مباشرة بواسطة التباينين s_1^2 و s_2^2 الملحوظين على العيّنة .

$$s_d = \sqrt{\frac{(104)^2}{327} + \frac{(118)^2}{286}}$$
$$= \sqrt{33{,}076\,5 + 48{,}685\,3} = \sqrt{81{,}761\,8} = 9{,}04 .$$

لنأخذ في هذا المثل درجة المعنوية $\alpha = 0{,}01$ ، يتناسب هذا الاحتمال مع القيمة :

$$t_{\alpha/2} = 2{,}58$$

من قيم المتغيّرة الطبيعية الممركزة المختصرة .

إذن فسحة قبول الفرضية H_0 ، التي تناسب درجة الاحتمال هذه ، هي :

$$-2,58 \times 9,04 < d < +2,58 \times 9,04$$

$$-23,32 < d < +23,32$$

يقع الفارق الملحوظ :

$$d = \overline{x_1} - \overline{x_2} = -30\,F$$

خارج هذه الفسحة : إذن هو فارق معنوي (كاشف) . يمكننا التأكيد ، دون فرص كثيرة في أن نخطىء (فرصة واحدة على 100) ، أنّه في هذا التجمّع السكّاني ينفق الموظّفون على المآكل أكثر من العمّال .

الفصل السابع

تنفيذ الأبحاث الإحصائية العشوائية

في بعض التطبيقات العملية ، يكون مجرّد الاستعمال البحت للبحث الإحصائي بدرجة واحدة مع احتمالات متساوية ، الذي عرضناه في الفصول السابقة باهظ الكلفة وقليل الفعالية . ويتضمّن وضع عملية الأبحاث الإحصائية موضع التنفيذ استعمال عدد معيّن من المناهج يتعلّق بعضها بطريقة تنظيم سحب العيّنة (تبسيط السحب ، تخفيض كلفة جمع المعلومات ، الخ . .) ويتعلّق البعض الآخر بتحسين فعالية الطريقة .

القسم I

تحديد العيّنة

1 . قاعدة البحث الإحصائي . ـ 2 . طرق سحب العيّنة : A . البحث الإحصائي النموذجي . استعمال جداول الأعداد العشوائية ؛ B . البحث الإحصائي المنهجي ؛ C . البحث الإحصائي بالعناقيد .ـ 3 . البحث الإحصائي مع احتمالات غير متساوية : A . المبدأ ؛ B . تطبيق سحب العيّنة عملياً ؛ C . الخصائص ؛ D . تحديد احتمالات السحب المثلى . ـ 4 . البحث الإحصائي على عدّة درجات : A . المبدأ ؛ B . الحسنات والسيئات ؛ C . الكيفيات العملية لحسب عيّنة على درجتين .

نفترض طريقة الأبحاث العشوائية أنّ لكلّ وحدة من المجتمع الإحصائي احتمالاً مختلفاً عن الصفر لأن تنتمي إلى العيّنة وأنّنا نعرف هذا الاحتمال . يقوم النهج الأكثر نموذجية على سحب الـ n وحدة ـ عيّنة باحتمالات متساوية من ضمن الـ N وحدة التي

تؤلّف المجتمع الإحصائي . هذه العملية تستدعي وجود قاعدة للبحث الإحصائي .

1 . قاعدة البحث الإحصائي

قاعدة البحث الإحصائي هي عبارة عن لائحة أو سجلّ بوحدات المجتمع الإحصائي دون حذف (لأنّه يجب أن يكون لكلّ وحدة احتمال مختلف عن الصفر لأن تعيّن) ودون تكرار (كي نضمن المساواة بين احتمالات الإخراج) .

من المهمّ بشكل خاص أن تكون قاعدة البحث الإحصائي كاملة وشاملة . في الواقع ، إذا كان السجلّ يتضمّن بعض التكرارات ، يسهل بشكل عام حذفها . وإذا اختفت ، لنقص في الاستيفاء اليومي ، بعض وحدات السجل ، يُلمس هذا الغياب حتماً عند القيام بالحملة . بالمقابل يجب أن نسعى لوضع لائحة على الأقلّ تقريبية بالوحدات الجديدة التي لم تدخل بعد في السجلّ ، وفي هذه اللائحة نقوم بأخذ عيّنة تأتي لتكمل العيّنة المأخوذة من قاعدة البحث الإحصائي الأصلية .

مثلاً . سجلّ شهادات السكن . لأجل حملاتها المتداولة حول الأسر ، تعتمد I.N.S.E.E [1] ، كقاعدة للبحث الإحصائي ، سجلّ شهادات السكن الناتج عن أحدث فرز سكّاني .

تحدّد الأسرة كمجموعة الأشخاص الذين يعيشون في مسكن واحد . والمساكن هي إمّا أمكنة إقامة رئيسية ، إمّا ثانوية إمّا أيضاً مساكن شاغرة . بناء على التعريف هناك توافق بين فكرة الأسرة وفكرة المسكن الرئيسي .

ضمن هذه الشروط ، تطرح على الباحثين القواعد التالية :

1 . عندما يكون أحد مساكن العيّنة ، عند تاريخ البحث ، المسكن الرئيسي لأسرة ما ، يجب أن نستجوب هذه الأسرة ، حتّى ولو لم تكن تشغل هذا المسكن عند تاريخ الفرز السكّاني .

لا يجب استبعاد المساكن الثانوية أو الشاغرة عند الفرز السكاني عن العيّنة : فهي قد تكون أصبحت مساكن رئيسية منذ هذا التاريخ وينبغي إذن زيارتها من قبل الباحثين .

2 . عندما يكون أحد مساكن العيّنة مسكناً ثانوياً عند تاريخ البحث لا يجب إجراء

Institut National de la Statistique et des Etudes Economiques (1)
المعهد الوطني للإحصاء والدراسات الإقتصادية .

المقابلة . في الواقع إذا شملت الحملة المساكن الثانوية ، فهذا قد يعطي الأسر التي تملك مسكناً ثانياً احتمالاً لأن تستجوب يبلغ ضعفي احتمال الأسر الأخرى .

من جهة أخرى ، ليس للمساكن « الجديدة » التي تمّ بناؤها بعد الفرز السكّاني الأخير ، أيّ فرصة لأن تعيّن بواسطة هذا النهج لأنّها لم تذكر في قاعدة البحث الإحصائي . يجب إذن أن نكمل هذه القاعدة بواسطة لائحة ، على الأقلّ تقريبية ، تتضمّن المساكن « الجديدة » : مثلاً ، لائحة برخصات البناء أو أيضاً سجلّ بالمساكن قيد التعمير . ونقوم بأخذ عيّنة متمّمة من هذه اللائحة تبعاً لنفس نسبة البحث الإحصائي كما في اللائحة الأصلية .

2 . طرق سحب العيّنة

إنّ سحب العيّنة هو عمليّة معقّدة ، لهذا نستعمل عملياً طرقاً عديدة (جداول الأعداد العشوائية ، السحوبات المنهجية ، السحوبات بالعناقيد أو الجماعات) لتبسيطه .

A . السحب النموذجي.. استعمال جداول الأعداد العشوائية

تقوم الطريقة النموذجية على سحب العيّنة مع إعطائنا لكلّ وحدة من المجتمع الإحصائي نفس الاحتمال لأن تُسحب كرفيقاتها . ولهذا يجب أن :

1 . نحصل على أو نضع قاعدة البحث الإحصائي ؛

2 . نرقّم الوحدات الإحصائية من 1 إلى N ؛

3 . نحدّد حجم العيّنة n ؛

4 . نسحب n عدداً محصوراً بين 1 وN ، مع إعطائنا لكلّ من الـ N رقم نفس احتمال السحب .

تشبه هذه العملية الأخيرة سحب n كرة من وعاء يحتوي N منها، مرقّمة من 1 الى N ولا نميّز بينها سوى بواسطة أرقامها . ويمكن إجراء السحوبات :

- إمّا مع ردّ الى الوعاء : سحوبات مستقلّة ،
- إمّا دون ردّ الى الوعاء : سحوبات مستنفِدة .

عملياً ، نعمد دوماً ، بشكل عام ، إلى السحوبات المستنفِدة . فهذه الطريقة تعطي في الواقع ، بالنسبة لعيّنتين بنفس المقدار ، تقديرات أدقّ وذلك لأنّ تباين عيّنة مستنفدة هو دوماً أصغر من تباين عيّنة مستقلّة (أنظر الفصل VI ، القسم I ، ص 247) .

أن نسحب بالصدفة ، وباحتمالات متساوية ، عيّنة من الوحدات في مجتمع إحصائي ما ليس بالأمر السهل كما قد يتبادر إلى الذهن بادىء الأمر . يجب أن يتحرّر العامل من أي تصوّر خلال اختياره وأن يتّبع لهذا الأمر نهجاً موضوعياً . وأبسط ما يخطر على البال هو أن نجري بسحب العيّنة كسحب اليانصيب ، بتسجيلنا الأرقام التي تعاين الوحدات الإحصائية على دواليب نجعلها تدور أو على أوراق مخلوطة داخل وعاء ، لكن فعالية هذه الطرق تصبح ضعيفة عندما يكون مقدار العيّنة كبيراً ـ عدّة آلاف أو أيضاً عدّة عشرات الآلاف من الوحدات الإحصائية . يمكننا عندئذٍ أن نستعمل جداول الأعداد العشوائية .

أ ـ وصف جدول الأعداد العشوائية

لقد وضع بعض الإحصائيين جداول تتضمّن سلاسل أرقام من 0 إلى 9 ، مسحوبة بالصدفة وباحتمالات متساوية . بحوزتنا إذن جداول Tipett ، جداول Fisher وYates ، جداول Kendall وBabington Smith ، جداول Burke Horton وجداول الـ Rand Corporation . وننقل في الملحق (الجدول 7) صفحة من جدول Kendall وBabington Smith .

إنّ هذه الجداول تسمح بتسهيل سحب العيّنة إلى حدّ بعيد .

ب ـ استعمال جول الأعداد العشوائية

مثلاً : لنفترض أنّه علينا سحب 9 وحدات من مجتمع إحصائي مؤلّف من 453 وحدة (معدّل أو نسبة البحث الإحصائي : t = 1/50) .

نحدّد بالصدفة المكان حيث سنبدأ بقراءة الجدول : مثلاً ، الألف الـ 36 ، السطر 11 ، العامود 13 من جدول Kendall وBabington Smith (أنظر الملحق : الجدول 7) . ثمّ نقرأ بالترتيب الأعمدة الثلاثة 13 ، 14 و15 من الأعلى إلى الأسفل (ويمكننا أيضاً أن نقرّر قراءة الجدول من أسفل إلى أعلى أو ، بالسطر ، من اليسار إلى اليمين ، الخ . .) . إذن العيّنة ستضمّن الوحدات التالية :

153, 358, 371, 126, 087, 262, 145, 421, 424

وقد استثنينا الأعداد 611 ، 960 ، 726 ، 723 ، 906 ، 936 ، 768 و970 لأنّها أكبر من 453 .

بعد ذلك نرتّب الأعداد التي حصلنا عليها :

087, 126, 145, 153, 262, 358, 371, 421, 424

مّما يسهّل البحث عن الوحدات المطابقة في السجلّ ويسمح ، في حالة السحوبات المستنفِدة ، باستبعاد الوحدات التي قد تعاين أكثر من مرّة .

إذا تمّ وضع قاعدة البحث الإحصائي على أداة معلوماتية ، يمكن تحديد العيّنة مباشرة بواسطة الحاسب الآلي الذي تزوّده بجدول أعداد عشوائية . وبالطبع لا يأخذ هذا النهج أهمّيته إلّا بالنسبة للعيّنات ذات الأحجام الكبيرة .

B . البحث الإحصائي المنهجي

إنّ طريقة السحوبات المنهجية تجنّبنا ضرورة سحب n عدداً بالصدفة . ومن ناحية أخرى ، يمكنها في بعض الحالات أن تظهر أكثر فعالية من الطريقة النموذجية .

أ ـ تعريف

تؤخذ وحدات العيّنة من المجتمع الاحصائي تبعاً لمتوالية حسابية نختار قاعدتها بالصدفة ونحسب أساسها بشكل يغطّي كامل المجتمع المرجع .

مثلاً . لنفترض أنّه علينا سحب عيّنة بنسبة 1/25 من مجتمع إحصائي مؤلّف من 453 وحدة .

نأخذ كقاعدة للمتوالية رقماً مسحوباً بالصدفة بين 1 و25 ، 17 مثلاً ، وكأساس لها الرقم 25 .

ستتضمّن العيّنة الوحدات ذات الرتب التالية :

17, 42, 67, 92, ..., 417, 442

ويصبح مقدار العيّنة مساوياً 19 إذا أعطانا السحب الأوّل كقاعدة رقماً محصوراً بين 1 و3 ، ومساوياً 18 إذا أعطانا السحب الأوّل رقماً بين 4 و25 .

ب ـ الخصائص

إنّ العيّنة التي نأخذها بواسطة سحب منهجي هي عيّنة عشوائية . إلّا أنّها توافق سحب عنقود أو جماعة واحدة مؤلّفة من كلّ الوحدات التي تنتمي أرقامها إلى ذات

المتوالية الحسابية . إذاً ، تكون دقّة النتائج مختلفة عن ما قد تؤول إليه الطريقة النموذجية .

لنأخذ مجتمعاً إحصائياً مؤلّفاً من N وحدة U_s يشار إليها بواسطة رقمها s :

$$s = 1, 2, \ldots, N$$

ونقتطع منه ، بواسطة السحب المنهجي ، عيّنة بنسبة البحث الإحصائي t=(1/k) سنفترض لتسهيل العرض أنّ N هي مضاعفة لـ k :

$$N = n \cdot k$$

حيث n هو مقدار العيّنة .

لنأخذ المتغيّرة X ، يمكننا ترتيب القيم X_s التي تأخذها هذه المتغيّرة بالنسبة لكل من وحدات المجتمع U_s في جدول له k سطراً و n عاموداً :

		1	2	3	...	j	...	n
	1	X_1	X_{1+k}	X_{1+2k}	...	$X_{1+(j-1)k}$	...	$X_{1+(n-1)k}$
	2	X_2	X_{2+k}	X_{2+2k}	...	$X_{2+(j-1)k}$	...	$X_{2+(n-1)k}$
	⋮	⋮	⋮	⋮		⋮		⋮
(العيّنة i) →	i	X_i	X_{i+k}	X_{i+2k}	...	$X_{i+(j-1)k}$	...	$X_{i+(n-1)k}$
	⋮	⋮	⋮	⋮		⋮		⋮
	k	X_k	X_{2k}	X_{3k}	...	X_{jk}	...	X_{nk}

تقوم طريقة السحوبات المنهجية على اختيار ، بالصدفة ، عدد بين 1 وk ، مثلاً i ، وعلى أن نأخذ في العيّنة الوحدات ذات الرتب i ، i+k ، i+2k ، الخ . . هذا النهج يعني إذن أن نسحب بالصدفة سطراً من الجدول السابق .

لنرمز بواسطة :

X_{ij} إلى قيمة المتغيّرة X بالنسبة للوحدة المرصوفة عند تقاطع السطر i مع العامود j ؛ $\overline{X_{i.}}$ إلى متوسّط X بالنسبة للسطر i :

$$X_{i.} = \frac{1}{n} \sum_{j=1}^{n} X_{ij} ;$$

$\bar{X}$ إلى المتوسّط العام للمجتمع الإحصائي :

$$\bar{X} = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n} X_{ij} = \frac{1}{k}\sum_{i=1}^{k}\frac{1}{n}\sum_{j=1}^{n} X_{ij} = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_{i.}\ ;$$

σ^2 إلى تباين المجتمع الإحصائي :

$$\sigma^2 = \frac{1}{nk}\sum_{i=1}^{k}\sum_{j=1}^{n}(X_{ij} - \bar{X})^2 .$$

		1	2	3	...	j	...	n	المتوسّطات
	1	X_{11}	X_{12}	X_{13}	...	X_{1j}	...	X_{1n}	$\bar{X}_{1.}$
	2	X_{21}	X_{22}	X_{23}	...	X_{2j}	...	X_{2n}	$\bar{X}_{2.}$
	⋮	⋮	⋮	⋮		⋮		⋮	⋮
العيّنة →	i	X_{i1}	X_{i2}	X_{i3}	...	X_{ij}	...	X_{in}	$\bar{X}_{i.}$
	⋮	⋮	⋮	⋮		⋮		⋮	⋮
	k	X_{k1}	X_{k2}	X_{k3}	...	X_{kj}	...	X_{kn}	$\bar{X}_{k.}$

بما أنّ السحب المنهجي يؤدّي إلى اختيار سطر بالصدفة مع احتمالات متساوية ، فإنّ $\bar{X}_{i.}$ هي متغيّرة عشوائية تأخذ القيم التالية :

$$\bar{X}_{1.}, \bar{X}_{2.}, \ldots, \bar{X}_{k.} .$$

مع الاحتمالات :

$$\frac{1}{k}, \frac{1}{k}, \ldots, \frac{1}{k} .$$

الأملان الرياضيان للمتوسّط والتردّد الملحوظين على العيّنة

بناء على تعريف الأمل الرياضي :

$$E(\bar{X}_{i.}) = \frac{1}{k}\sum_{i=1}^{k}\bar{X}_{i.} = \bar{X} .$$

إنّ متوسّط عينة منهجية هو مقدّر غير متحيّز لمتوسّط المجتمع الإحصائي .

يمكننا بسط هذه النتيجة إلى تقدير تردّد خاصّة معيّنة في المجتمع الإحصائي ، باعتبارنا المتغيّرات X_{ij} متغيّرات برنولي (أنظر الفصل II ، القسم 1 ، ص 72) تأخذ القيمة 1 عندما تملك الوحدة المأخوذة هذه الخاصّة ، والقيمة صفر عندما لا تملكها .

بالنسبة للسطر i :

$$f_{i} = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$$ (متوسّط X_{ij} في العيّنة) .

و :

$$E(f_i) = \frac{1}{k} \sum_{i=1}^{k} f_i = p$$

حيث p تمثّل نسبة الوحدات التي تملك الخاصّة في مجمل المجتمع الإحصائي .

إنّ تردّد خاصّة في عيّنة منهجية هو مقدّر غير متحيّز لنسبة الوحدات التي تملك هذه الخاصّة في المجتمع الإحصائي .

تباين متوسّط العيّنة

بناء على التعريف :

$$V(\overline{X}_{i.}) = \frac{1}{k} \sum_{i=1}^{k} (\overline{X}_{i.} - \overline{X})^2 .$$

وإذا استبدلنا $\overline{X}_i$ بعبارتها :

$$\begin{aligned} V(\overline{X}_{i.}) &= \frac{1}{k} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} \frac{X_{ij}}{n} - \overline{X} \right)^2 \\ &= \frac{1}{k} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} \frac{X_{ij} - \overline{X}}{n} \right)^2 \\ &= \frac{1}{n^2 k} \sum_{i=1}^{k} \left[\sum_{j=1}^{n} (X_{ij} - \overline{X}) \right]^2 \end{aligned}$$

إذا وسّعنا المربّع ، ننتهي الى :

$$\begin{aligned} V(\overline{X}_{i.}) &= \frac{1}{n^2 k} \sum_{i=1}^{k} \left[\sum_{j=1}^{n} (X_{ij} - \overline{X})^2 + 2 \sum_{j<j'=1} \sum^{n} (X_{ij} - \overline{X})(X_{ij'} - \overline{X}) \right] \\ &= \frac{1}{n^2 k} \sum_{i=1}^{k} \sum_{j=1}^{n} (X_{ij} - \overline{X})^2 + \frac{2}{n^2 k} \sum_{i=1}^{k} \sum_{j<j'=1} \sum^{n} (X_{ij} - \overline{X})(X_{ij'} - \overline{X}) . \end{aligned}$$

إنّ العنصر الأوّل يساوي تباين متوسّط عيّنة بنفس الحجم مسحوبة بواسطة الطريقة النموذجية[1] :

$$\frac{\sigma^2}{n} = \frac{1}{n} \cdot \frac{1}{nk} \sum_{i} \sum_{j} (X_{ij} - \overline{X})^2 .$$

(1) بفارق مُعامل الاستنفاد ، بما أنّ سحب العيّنة المنهجية يتمّ ، بناء على التعريف ، دون ردّ (راجع الفصل VI ، ص 247) .

إذاً ، يكون البحث الإحصائي المنهجي أكثر أو أقلّ فعالية من البحث الإحصائي النموذجي حسب إشارة العنصر الثاني ، ويمكننا كتابته :

$$\frac{2}{n^2}\sum_{j<j'}\sum\left[\frac{1}{k}\sum_{i}(X_{ij}-\bar{X})(X_{ij'}-\bar{X})\right].$$

إنّ الكمية بين رمزي التعانق [ ] هي بالمعنى الواسع التغاير بين العناصر المتوافقة في العامودين j وَ'j من جدول X_{ij} [1] . بالتالي ، يمثّل العنصر الثاني من المجموع ، بفارق عامل ضرب ، متوسّط التغاير بين كلّ العواميد .

- إذا كان انتشار الوحدات في قاعدة البحث الإحصائي قد تمّ بالصدفة ، فإنّ متوسّط التغاير بين العواميد يساوي صفراً . في هذه الحالة ، تباين البحث الإحصائي المنهجي هو نفس تباين البحث الإحصائي النموذجي .
- إذا كان متوسّط التغاير بين الأعمدة سلبياً ، مثلاً لأنّ الوحدات القريبة من بعضها في القاعدة تتشابه فيما تكون الوحدات المتباعدة ، بشكل عام ، مختلفة عن بعضها البعض ، فإنّ تباين البحث الإحصائي المنهجي هو أصغر من تباين البحث الاحصائي النموذجي .
- إذا كان متوسّط التغاير بين الأعمدة إيجابياً ، فإنّ تباين البحث الإحصائي المنهجي هو أكبر من تباين البحث الإحصائي النموذجي . والحالة القصوى هي حيث تكون المتغيّرة X دورية ، بدورة k :

$$X_i = X_{i-k} = X_{i+2k} = \dots .$$

عندها يكون تباين التقدير حدّاً أقصى .

بالمختصر ، تكون دقّة البحث الإحصائي المنهجي أكبر بشكل عام من دقّة البحث الإحصائي العادي ذي الحجم نفسه . بعبارة أدقّ :

- إذا كان بالإمكان اعتبار ترتيب الوحدات الإحصائية ، في السجلّ المعتمد كقاعدة

---

(1) عبارة التغاير الحقيقية بين العامودين j وَj' هي :

$$\frac{1}{k}\sum_{i}(X_{ij}-\bar{X}_{.j})(X_{ij'}-\bar{X}_{.j'}),$$

حيث $\bar{X}_{.j}$ و $\bar{X}_{.j'}$ تشيران على التوالي إلى متوسّط المتغيّرة X في العامود j والعامود j' .

لدينا :

$$\frac{1}{k}\sum_{i}(X_{ij}-\bar{X})(X_{ij'}-\bar{X})=\frac{1}{k}\sum_{i}(X_{ij}-\bar{X}_{.j})(X_{ij'}-\bar{X}_{.j'})+(\bar{X}_{.j}-\bar{X})(\bar{X}_{.j'}-\bar{X}).$$

للبحث الإحصائي عشوائياً فإنّ طريقتي البحث متعادلتان .

- إذا كان يوجد بين الوحدات التي تشغل رتباً متجاورة في السجل عناصر شبه ، فإنّ دقّة البحث الإحصائي المنهجي هي أفضل .

وغالباً ما يكون الأمر على هذا النحو على الصعيد العملي .

مثلاً . لأسباب تتعلّق بالسرعة وبالكلفة ، يتمّ فرز الإحصاء السكّاني الفرنسي على عيّنة بنسبة 1/20 ، وتؤخذ هذه العيّنة بواسطة سحب منهجي من شهادات السكن . وبما أنّه يتمّ ترتيب هذا السجلّ على أساس الشوارع ، الأحياء ، البلدات والمناطق ، فإنّ طريقة السحب هذه تضمن توزيعاً جغرافياً مرضياً للعيّنة . بالنسبة للعديد من الخصائص الاجتماعية - الاقتصادية (الفئة الإجتماعية - المهنية ، النشاط الإقتصادي ، الخ . .) التي تكون على علاقة وثيقة مع مكان الإقامة ، نحصل بهذه الطريقة على دقة كبيرة جدّاً بالنسبة لما قد يعطيه البحث الإحصائي النموذجي .

- بالمقابل ، إذا تحكّمت أيّ دورية بترتيب الوحدات في السجل ، قد تؤدّي الطريقة هذه إلى أخطاء فادحة في التقدير ، خاصّة إذا كانت الدورة مضاعفاً ثانوياً لأساس متوالية السحب الحسابية . ولحسن الحظ قليلاً ما نصادف هذه الحالة .

C . البحث الإحصائي بالعناقيد أو بالجماعات

أ - التعريف

إنّ البحث الإحصائي بالعناقيد يختلف عن البحث الإحصائي النموذجي بكوننا لا نسحب وحدات العيّنة واحدة واحدة ، بل « برزم » ندعوها عناقيد أو جماعات .

يتألّف العنقود إذن من مجموعة وحدات إحصائية ، وكلّ وحدة تتعلّق بعنقود واحد فقط .

هكذا ، فالأسرة ، أي مجموعة الأشخاص الذين يقطنون مسكناً واحداً ، هي عنقود من الأفراد ، والبناية هي عنقود من المساكن أي من الأسر ، المؤسّسة هي عنقود من الموظّفين ، الخ .

ب - الخصائص

إنّ السحب بالعناقيد يسهّل وضع قاعدة البحث الإحصائي : من الأسهل مثلاً

وضع لائحة مساكن بدلاً من لائحة أشخاص ، وضع سجلّ بالمؤسّسات بدلاً من سجل بالموظّفين .

إلاّ أنّ تبريره يكمن بشكل خاص في تخفيض كلفة تحقيق البحث على أرض الدراسة . وبما أن الوحدات التي تؤلّف العنقود الواحد تكون متجاورة بشكل عام ، فإن السحب بالعناقيد يسمح بتوفير جوهري في نفقات التنقّل بالنسبة لنفس عدد الوحدات موضع الفحص .

بالمقابل، غالباً ما تكون الوحدات الإحصائية التي تؤلّف نفس العنقود متشابهة. إذن لا يمكن تشبيه العيّنة المأخوذة بهذه الطريقة بعيّنة نموذجية بنفس الحجم : أكثر الأحيان يعطي البحث الإحصائي بالعناقيد تقديرات أقلّ دقّة من بحث إحصائي نموذجي بنفس الحجم . مع ذلك ، وعند كلفة ثابتة ، تلعب المقارنة دوراً لصالح السحب بالعناقيد .

لنأخذ مجتمعاً إحصائياً مؤلّفاً من N وحدة ، ولنفترض ، لتسهيل العرض، أنّه مؤلّف من M عنقود بنفس الحجم يحتوي كلّ منها على h وحدة :

$$N = M.h$$

تحتوي العيّنة على m عنقود ، ومقدارها هو :

$$n = m.h$$

لنأخذ المتغيّرة X ، يمكننا ترتيب القيم التي تأخذها هذه المتغيّرة بالنسبة لكلّ من الوحدات الإحصائية في جدول له M سطراً و h عاموداً يشبه الجدول الذي استعملناه لتحليل البحث الإحصائي المنهجي :

رقم العنقود	1	2	...	j	...	h	المتوسّطات
1	X_{11}	X_{12}	...	X_{1j}	...	X_{1h}	$\overline{X}_1$
2	X_{21}	X_{22}	...	X_{2j}	...	X_{2h}	$\overline{X}_2$
⋮	⋮	⋮		⋮		⋮	⋮
i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ih}	$\overline{X}_i$
⋮	⋮	⋮		⋮		⋮	⋮
M	X_{M1}	X_{M2}	...	X_{Mj}	...	X_{Mh}	$\overline{X}_M$

كلّ سطر من الجدول هو عبارة عن عنقود . يقوم البحث الإحصائي بالعناقيد على أن نسحب بالصدفة ، ودون ردّ بشكل عام ، عيّنة تتكوّن من m سطراً

نشير إلى صلة القرابة ، من الناحية الشكلية ، بين البحث الإحصائي بالعناقيد والبحث الإحصائي المنهجي حيث لا نسحب سوى سطر واحد . مما يسمح لنا ، عندما تكون العناقيد متساوية ، بتعميم النتائج التي حصلنا عليها بالنسبة للبحث المنهجي .

لنرمز بواسطة :

$\overline{X}_i$ إلى متوسّط X بالنسبة للسطر i :

$$\overline{X}_i = \frac{1}{h} \sum_{j=1}^{h} X_{ij}.$$

$\overline{X}$ إلى المتوسّط العام للمجتمع الإحصائي :

$$\overline{X} = \frac{1}{Mh} \sum_{i=1}^{M} \sum_{j=1}^{h} X_{ij} = \frac{1}{M} \sum_{i=1}^{M} \frac{1}{h} \sum_{j=1}^{h} X_{ij} = \frac{1}{M} \sum_{i=1}^{M} \overline{X}_i.$$

$\overline{x}$ إلى متوسّط X بالنسبة للعيّنة :

$$\overline{x} = \frac{1}{mh} \sum_{i=1}^{m} \sum_{j=1}^{h} X_{ij} = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{h} \sum_{j=1}^{h} X_{ij} = \frac{1}{m} \sum_{i=1}^{m} \overline{X}_i.$$

وبما أنّ البحث الإحصائي بالعناقيد يعني أن نسحب بالصدفة ، مع أو بدون ردّ إلى الوعاء ، m سطرا من ضمن M ، فإنّ $\overline{X}_i$ هي متغيّرات عشوائية ، يمكنها أن تأخذ القيم التالية :

$$\overline{X}_1, \overline{X}_2, \ldots, \overline{X}_M .$$

مقدّر متوسّط المجتمع الإحصائي

يقدّر متوسّط المجتمع الإحصائي $\overline{X}$ بواسطة متوسّط العيّنة $\overline{x}$. بالفعل :

$$\overline{x} = \frac{1}{m} \sum_{i=1}^{m} \overline{X}_i$$

وبناء على خصائص الأمل الرياضي (الفصل I ، ص 56) :

$$E\{\overline{x}\} = \frac{1}{m} \sum_{i=1}^{m} E\{\overline{X}_i\}.$$

ونعرف أنّه في حال أجري سحب العناقيد مع أو بدون ردّ :

$$E\{\overline{X}_i\} = \overline{X}.$$

بالتالي :

$$E\{\overline{x}\} = \overline{X}.$$

متوسّـط العيّـنة هو مقدّر غير متحيّـز لمتوسّـط المجتمع الإحصائي .

تباين متوسّـط العيّـنة

يمكننا اعتبار $\overline{x}$ كمتوسّـط العيّـنة المؤلّـفة من المتوسّـطات $\overline{X_i}$ للأسطر m المعيّـنة بالقرعة . وبفضل النتائج المتعلّـقة بتباين متوسّـط العيّـنة (الفصل VI ، القسم I ، ص 244) :

في حالة السحوبات المستقلّـة : $- V(\overline{x}) = \frac{V(\overline{X_i})}{m}$

في حالة السحوبات المستنفِدة : $- V(\overline{x}) = \frac{V(\overline{X_i})}{m} \frac{M-m}{M-1}$

حيث ترمز $V(\overline{X_i})$ إلى تباين متوسّـطات الأسطر في الجدول والتي حسبناها سابقاً حول موضوع البحث الأحصائي المنهجي . وتقودنا مقارنة فعالية بحث بالعناقيد مع فعاليةبحث نموذجي إلى النتائج كما في حالةالبحث المنهجي : يكون البحث بالعناقيد أقلّ فعالية من البحث النموذجي ذي الحجم نفسه عندما تكون الوحدات التي تؤلّـف العناقيد متشابهة .

فعالية العناقيد

عندما يكون الخيار ملكنا ، من الأفضل :

- أن لا تكون العناقيد ضخمة جدّاً ، بشكل يكون فيه عددها كافياً ؛
- أن تكون أحجامها متماثلة قدر الإمكان ؛
- أن تكون الوحدات التي تؤلّـفها غير متجانسة قدر الإمكان من ناحية الخاصّـة موضع الدراسة . عندها نقول أنّ العناقيد فعّـالة .

وقد يكون القطع فعّـالاً بالنسبة لدراسات معيّـنة ، وغير فعّـال بالنسبة لدراسات أخرى .

فالبناية مثلاً هي عنقود فعّـال نسبياً لتقدير توزيع السكّـان حسب الجنس ، العمر ، العمل أو البطالة ؛ وهي عنقود غير فعّـال بالنسبة لدراسة بواسطة الفئة الاجتماعية ـ المهنية .

غالباً ما تكون الأسرة عنقوداً غير فعّـال ، وذلك لأنّ أعضاءها يميلون ، من علّـة وجهات نظر ، إلى أن يتشابهوا . والأمر يكون كذلك بصورة خاصّـة بالنسبة للدراسة حول قراءة الصحف ، حول العطل ، حول الآراء السياسية .

البحث الإحصائي المساحي

البحث الإحصائي المساحي هو نوع خاص من الأبحاث بالعناقيد : إذ يتألّف كلّ عنقود من مساحة معيّنة بواسطة حدود يسهل التعرّف إليها : شوارع ، طرقات ، مجاري مياه ، الخ . .

هكذا ، يتمّ تقطيع مجمل الأرض الخاضعة للدراسة إلى مساحات وتتعلّق كلّ وحدة إحصائية (شخص ، أسرة ، مؤسّسة صناعية) بمساحة واحدة فقط .

من حسنات هذه الطريقة أنّها لا تستدعي عملية استيفاء يومي لقاعدة البحث الاحصائي كما بالنسبة لعيّنة من المساكن أو المؤسّسات .

وسيئاتها هي :

- من جهة ، عدم فعالية المساحة كعنقود : غالباً ما تميل الوحدات الإحصائية المتجاورة جغرافياً إلى التشابه ؛
- عملياً ، صعوبة تحديد مساحات تتضمّن تقريباً نفس عدد الوحدات الإحصائية وصغيرة بشكل كاف وذات حدود يسهل التعرّف إليها .

3 . البحث الإحصائي باحتمالات غير متساوية

إفترضنا إلى الآن أنّ سحب العيّنة يتمّ باحتمالات متساوية . لنأخذ هذه المرّة سحباً تكون فيه لوحدات المجتمع الإحصائي فرص مختلفة في التعيين .

A . المبدأ

لنأخذ مجتمعاً إحصائياً مؤلّفاً من N وحدة U_i . يقوم البحث الإحصائي باحتمالات غير متساوية على أن نعطي لكلّ من الوحدات :

$$U_1, U_2, \ldots, U_s, \ldots, U_N,$$

احتمالات غير متساوية ، ولكن معروفة ومختلفة عن الصفر ، في أن تنتمي للعيّنة :

$$p_1, p_2, \ldots, p_s, \ldots, p_N ;$$

$$\sum_{s=1}^{N} p_s = 1 , \qquad p_s \neq 0 \ \forall s \qquad \text{(مهما تكن s)}$$

B . تحقيق سحب العيّنة عملياً

على الصعيد العملي ، يجري سحب العيّنة باحتمالات غير متساوية بواسطة

طريقة الحواصل المتراكمة أو المجمّعة .

مثلاً . لنفترض أنّنا نريد أن نسحب بالصدفة مؤسّستين صناعيتين من مجتمع إحصائي يتكوّن من ست مؤسّسات ، وذلك باحتمالات تناسبية مع عدد موظّفي كلّ مؤسّسة .

نحسب عدد الموظّفين المتراكم :

المؤسّسة رقم	عدد الموظفين	عدد الموظفين المتراكم
1	1200	1200
2	300	1500
3	1800	3300
4	720	4020
5	600	4620
6	1380	6000
المجموع أو الحاصل	6000	

السحوبات مستقلّة

نسحب بالصدفة ، مثلاً في جدول أعداد عشوائية ، عدداً من 4 أرقام محصوراً بين 0000 و9999 .

إذا كان هذا العدد محصوراً بين 0000 و1200 ، نأخذ المؤسّسة رقم 1 .
إذا كان هذا العدد محصوراً بين 1201 و1500 ، نأخذ المؤسّسة رقم 2 .
إذا كان هذا العدد محصوراً بين 1501 و3300 ، نأخذ المؤسّسة رقم 3 .
إذا كان هذا العدد محصوراً بين 3301 و4020 ، نأخذ المؤسّسة رقم 4 .
إذا كان هذا العدد محصوراً بين 4021 و4620 ، نأخذ المؤسسة رقم 5 .

إذا كان هذا العدد محصوراً بين 4621 و6000 ، نأخذ المؤسّسة رقم 6 .

وإذا كان أكبر من 6000 ، نعيد السحب حتى نحصل على عدد أصغر من أو يساوي 6000 .

ونعيد العملية من أجل سحب المؤسّسة الثانية ، قد يحصل إذن أن نعيّن نفس المؤسّسة مرّتين .

السحوبات مستنفدة

تُسحب المؤسّسة الأولى بالطريقة المشار إليها أعلاه . ولكن ، عند السحب الثاني ، تُرفع المؤسّسة المسحوبة سابقاً من الوعاء . إذن تتغيّر احتمالات خروج مؤسّسة معيّنة من سحب لآخر .

عملياً ، نعمد بشكل عام إلى سحوبات منهجية : في مثلنا ، نأخذ كقاعدة لمتوالية السحب الحسابية ، عدداً نختاره بالصدفة بين 0 و3000 ، 1584 مثلاً ، ونأخذكاساس لها 3000 . العددان المسحوبان إذن هما 1584 و4584 اللذان يشيران على التوالي إلى المؤسّستين رقم 3 و5 .

C . الخصائص

عندما تكون الوحدات الإحصائية ذات أحجام مختلفة (مؤسّسات صناعية ، تجمّعات سكنية ، بلدات ، الخ . .) فإنّ تحديد وحدات العيّنة باحتمالات غير متساوية ، تقريباً تناسبية مع أحجامها ، يسمح بتحسين دقّة التقديرات .

بالمقابل ، لا يعود بالإمكان فرز العيّنة كالإحصاء السكّاني ، وذلك لأنّه يجب ترجيح المشاهدات المستقلّة بمعكوس احتمالات السحب .

لنأخذ المتغيّرة X ، ونعود إلى رموزنا المعتادة (الفصل السادس ، ص 241) ، سوف نشير :

- في المجتمع الإحصائي ،

بواسطة X_s إلى قيمة المتغيّرة X بالنسبة للوحدة U_s ، $s = 1, 2, \ldots, N$ ؛

بواسطة m إلى متوسّط X : $m = \frac{1}{N} \sum_{s=1}^{N} X_s$

- في العيّنة ،

بواسطة x_i إلى قيمة المتغيّرة X بالنسبة لوحدة العيّنة U_i ، $i = 1, 2, \ldots, n$

أ - مقدّر متوسّط المجتمع الإحصائي

يُقدّر متوسّط المجتمع الإحصائي m ، ليس بواسطة متوسّطة العيّنة $\overline{x}$ كما في حالة البحث الإحصائي باحتمالات متساوية ، ولكن بواسطة :

$$\overline{x}' = \frac{1}{N} \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{x_i}{p_i}.$$

في الواقع ، بفضل خصائص الأمل الرياضي (الفصل I ، ص 55) :

$$E(\overline{X}) = E\left\{\frac{1}{N}\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{p_i}\right\} = \frac{1}{N}\cdot\frac{1}{n}\sum_{i=1}^{n}E\left\{\frac{x_i}{p_i}\right\}$$

لكن ، إذا اعتبرنا أن سحب العيّنة قد تمّ مع ردّ ، وبناء على تعريف الأمل الرياضي :

$$E\left\{\frac{x_i}{p_i}\right\} = \sum_{s=1}^{N} p_s\frac{X_s}{p_s} = \sum_{s=1}^{N} X_s = Nm.$$

بالتالي :

$$E(\overline{X}) = \frac{1}{N}\cdot\frac{1}{n}\sum_{i=1}^{n} Nm = \frac{1}{N}\cdot\frac{1}{n}\cdot nNm = m.$$

في حالة: عيّنة مسحوبة مع ردّ ، $\overline{X}$ هو مقدّر غير متحيّز لمتوسّط المجتمع الإحصائي .

بالمقابل ، إذا كانت السحوبات مستنفِدة ، فإنّ احتمالات الخروج p_s بالنسبة لكلّ وحدة إحصائية تتغيّر من سحب لآخر ، ممّا يجعل $\overline{X}$ مقدّراً متحيّزاً لـ m ، إلاّ أنّ هذا التحيّز يكون عملياً بشكل عام دون أهمّية .

ب ـ تباين المقدّر

سوف نفترض أنّ السحوبات قد جرت مع ردّ :

$$V(\overline{X}) = V\left\{\frac{1}{N}\cdot\frac{1}{n}\sum_{i=1}^{n}\frac{x_i}{p_i}\right\} = \frac{1}{N^2}\cdot\frac{1}{n^2}V\left\{\sum_{i=1}^{n}\frac{x_i}{p_i}\right\}$$
$$= \frac{1}{N^2}\cdot\frac{1}{n^2}\sum_{i=1}^{n}V\left\{\frac{x_i}{p_i}\right\} = \frac{1}{N^2}\cdot\frac{1}{n^2}\cdot nV\left\{\frac{x_i}{p_i}\right\}.$$

وذلك بفضل خصائص التباين (الفصل I ، ص 61) ، وحيث القيم x_i/p_i هي متغيّرات عشوائية مستقلّة .

من جهة أخرى وبناء على تعريف التباين :

$$V\left\{\frac{x_i}{p_i}\right\} = E\left\{\left(\frac{x_i}{p_i} - E\left\{\frac{x_i}{p_i}\right\}\right)^2\right\} = E\left\{\left(\frac{x_i}{p_i} - Nm\right)^2\right\}.$$

وإذا استبدلنا الأمل الرياضي بعبارته ، نحصل على :

$$V\left\{\frac{x_i}{p_i}\right\} = \sum_{s=1}^{N} p_s\left(\frac{X_s}{p_s} - Nm\right)^2$$

أو ، من خلال قاعدة التباين المتبسّطة (الفصل I ، ص 63) :

$$V\left\{\frac{x_i}{p_i}\right\} = \sum_{s=1}^{N} \frac{X_s^2}{p_s} - N^2 m^2 .$$

بالتالي :

$$V(\bar{x}) = \frac{1}{N^2} \cdot \frac{1}{n} \sum_{s=1}^{N} \frac{X_s^2}{p_s} - \frac{m^2}{n} .$$

D . تحديد احتمالات السحب المثلى

كيف نختار احتمالات السحب p_s كي نحصل على أفضل تقدير ممكن ؟
المقصود هو ، على وجه الدقّة ، أن نحدّد قيم الاحتمالات

$$p_1, p_2, \ldots, p_s, \ldots, p_N$$

التي تجعل من تباين المقدّر $V(\bar{x})$ حدّاً أدنى ، وتربط بين هذه الاحتمالات العلاقة التالية :

$$\sum_{s=1}^{N} p_s = 1 .$$

إنّها إذن مسألة حدّ أدنى مرتبط .

تذكير رياضيات : الحدّ الأقصى المرتبط لدالّة معيّنة

لنأخذ الدالة $f(x_1, x_2, \ldots, x_n)$ بـ n متغيّرة $x_1, x_2, \ldots, x_n$ تحقّق في ما بينها العلاقة التالية :

$$h(x_1, x_2, \ldots, x_n) = k .$$

نحصل على حدّ الدالّة $f(x_1, x_2, \ldots, x_n)$ الأقصى (الأدنى أو الأعلى) المرتبط بالعلاقة .

$$h(x_1, x_2, \ldots, x_n) = k$$

إذا وجدنا الحدّ الأقصى للعبارة :

$$g(x_1, x_2, \ldots, x_n) = f(x_1, x_2, \ldots, x_n) + \lambda[h(x_1, x_2, \ldots, x_n) - k]$$

حيث λ هي متغيّر وسيطي نسمّيه مضروب لاغرانج (Lagrange) .
إنّ الـ n + 1 علاقة التالية :

$$\partial g/\partial x_1 = 0 \qquad \text{(تفاضل g بالنسبة لـ } x_1 = \text{ صفر)}$$
$$\partial g/\partial x_2 = 0$$
$$\vdots$$
$$\partial g/\partial x_n = 0$$
$$h(x_1, x_2, \ldots, x_n) = k$$

تسمح بتحديد قيم $x_1, x_2, \ldots, x_n$ التي تجعل $f(x_1, x_2, \ldots, x_n)$ حداً أقصى ، وكذلك قيمة

إذن ، كي نحدّد القيم $p_1, p_2, \ldots, p_s, \ldots, p_N$ التي تجعل $V(\bar{x}')$ حداً أدنى ، مع الشرط :

$$\sum_{s=1}^{N} p_s = 1$$

سوف نبحث عن الحدّ الأدنى للعبارة التالية :

$$W(p_1, \ldots, p_N) = V(\bar{x}') + \lambda\left(\sum_{s=1}^{N} p_s - 1\right)$$

نحصل على القيم p_s التي تناسب هذا الحدّ الأدنى إذا صفّرنا الـ N مشتقّة جزئية :

$$\frac{\partial W}{\partial p_s} = \frac{1}{N^2}\cdot\frac{1}{n}\left(-\frac{X_s^2}{p_s^2}\right) + \lambda = 0 \qquad s = 1, 2, \ldots, N$$

إذن :

$$X_s/p_s = \sqrt{N^2 n\lambda}$$

يمكننا إذن أن نكتب :

$$\frac{X_1}{p_1} = \frac{X_2}{p_2} = \cdots = \frac{X_N}{p_N} = \frac{X_1 + X_2 + \cdots + X_N}{p_1 + p_2 + \cdots + p_N} = \sum_{s=1}^{N} X_s = Nm$$

بحكم الشرط :

$$\sum_{s=1}^{N} p_s = 1$$

بالتالي ، يجب اختيار احتمال تعيين الوحدة U_s بشكل :

$$p_s = \frac{X_s}{\sum_{s=1}^{N} X_s}$$

في الحقيقة ، لا يمكن تحديد قيم p_s على وجه الدقّة ، فهذا التحديد يفترض معرفة كاملة لمقاييس المجتمع الإحصائي بالنسبة للمتغيّرة المدروسة . وفي هذه الحالة ، لا يعود مقدّر المتوسّط متغيّرة عشوائية ، ولكن يصبح عدداً ثابتاً وتباينه يساوي صفراً ، كما يمكننا أن نستنتج :

$$\begin{aligned} V(\bar{x}) &= \frac{1}{N^2}\cdot\frac{1}{n}\sum_{s=1}^{N}\frac{X_s^2}{p_s} - \frac{m^2}{n} \\ &= \frac{1}{N^2}\cdot\frac{1}{n}\sum_{s=1}^{N} X_s\cdot Nm - \frac{m^2}{n} \\ &= \frac{1}{N^2}\cdot\frac{1}{n}\cdot Nm\cdot Nm - \frac{m^2}{n} = 0 \end{aligned}$$

على الصعيد العملي ، نعطي لكلّ وحدة إحصائية احتمال خروج يتناسب مع « حجمها » (عدد السكّان في مجمّع سكني معيّن ، عدد الموظّفين في مؤسّسة ما ، الخ . .) ، كون هذا الحجم ، بشكل عام ، يتناسب تقريباً مع المتغيّرات الكمّية التي قد تهمّنا دراستها .

4 . البحث الإحصائي على عدّة درجات

A . المبدأ

يقوم البحث الإحصائي على عدّة درجات على تعيين وحدات العيّنة بالتسلسل :

- عند درجة السحب الأولى ، نختار بالصدفة عيّنة من الوحدات الأوّلية ؛
- عند درجة السحب الثانية ، في كلّ وحدة عيّنة أوّلية نسحب عيّنة من الوحدات الثانوية ؛
- عند درجة السحب الثالثة ، نسحب في كلّ وحدة عيّنة ثانوية ، عيّنة من الوحدات الثلثية ، الخ . .

بالطبع ، يجب أن تكون كلّ وحدة ثانوية متعلّقة بوحدة أوّلية واحدة فقط ، وكلّ وحدة ثلثية متعلّقة بوحدة ثانوية واحدة فقط ، الخ . .

مثلاً . لنفترض أنّنا بحاجة إلى تعيين عيّنة من الأراضي الزراعية .

بدلاً من أن نضع لائحة شاملة للأراضي الزراعية الموجودة في كلّ البلد وأن نسحب مباشرة عيّنة منها ، يمكننا :

- عند الدرجة الأولى ، أن نسحب عيّنة من المقاطعات تشكّل الوحدات الأوّلية ؛

- عند الدرجة الثانية ، وفي كلّ من المقاطعات المأخوذة ، أن نسحب عيّنة من البلدات (الوحدات الثانوية) ؛

- في كلّ من بلدات العيّنة ، أن نضع لائحة كاملة بالأراضي الزراعية ونسحب ، عند الدرجة الثالثة ، عيّنة منها (الوحدات الثلثية) .

B . الحسنات والسيئات

يسمح البحث الإحصائي على عدّة درجات بتسهيل وضع قاعدة البحث الإحصائي : يكفي مثلاً أن نضع لائحة بالأراضي الزراعية بالنسبة لبلدات العيّنة فقط . نتجنّب بهذه الطريقة ضرورة وضعها لمجمل البلد .

ولكن ، كما بالنسبة للبحث الإحصائي بالعناقيد ، تكمن الفائدة الحقيقية في تخفيض كلفة الحملة بالنسبة لنفس العدد من الوحدات المدروسة: إذ يؤمّن البحث الإحصائي بعدّة درجات حصراً جغرافياً للوحدات موضع المراقبة يسمح بتخفيض نفقات النقل إلى حدّ بعيد

بالمقابل ، عادة ما تكون دقّة التقديرات بالنسبة لعيّنة مسحوبة على عدّة درجات أقلّ جودة من دقّتها بالنسبة لعيّنة نموذجية بنفس المقدار : الأراضي الزراعية التي تنتمي إلى نفس البلدة تميل أكثر الأحيان إلى التشابه ؛ « فعل العنقود » هو غالباً غير ملائم .

إلاّ أنّه ، عند كلفة واحدة ، تتغلّب فعالية بحث إحصائي بعدّة درجات على فعالية بحث إحصائي بدرجة واحدة . في الواقع ، تُجرى الحملات الرئيسية إنطلاقاً من خطّة بحث إحصائي بعدّة درجات . بصورة خاصّة ، تعتمد الـ I.N.S.E.E لحملاتها حول الأفراد خطّة بحث إحصائي بثلاث درجات : مقاطعة ، بلدة أو تجمّع سكّاني ، مسكن .

من أجل وضع خطّة البحث الإحصائي ، المسألة الأساسية هي مسألة التوزيع الأمثل للعيّنة بين الوحدات الأوّلية والوحدات الثانوية . في الواقع ، يمكننا مثلاً ، دون أن نغيّر كلفة الحملة ، أن نزيد من عدد وحدات العيّنة الأوّلية على أن ننقص بالتلازم عدد وحدات العيّنة الثانوية في كلّ وحدة أوّلية وحتّى أن ننقص من العدد الإجمالي للوحدات الثانوية .

كي نسهّل العرض ، سوف نقتصر فيما يلي على معالجة البحث الإحصائي بدرجتين

C . الكيفيات العملية لسحب عيّنة على درجتين

مبدئياً ، يكون اختيار عدد وحدات العيّنة الأوّلية وعدد وحدات العيّنة الثانوية في كلّ وحدة أوّلية اختياراً حرّاً بالكامل .

إلاّ أنّه على الصعيد العملي من الأفضل الحصول على عيّنة يمكننا تعدادها كما في طريقة الفرز ، أي بعبارة أخرى دون أن يكون من الضروري إعطاء ترجيحات مختلفة لمختلف المشاهدات الفردية المستقلّة : عندئذٍ يُقدّر متوسّط المجتمع الإحصائي الكلّي بالمتوسّط المناسب المحسوب على العيّنة ، وتقدّر النسبة بالتردّد المناسب الملحوظ على العيّنة . لهذا من الضروري أن يكون لكلّ وحدة (ثانوية) من المجتمع الإحصائي ، بحكم مختلف درجات البحث الإحصائي ، نفس احتمال الانتماء إلى العيّنة . ونقول أنّ هذه العيّنة هي مرجّحة بذاتها .

هناك طريقتان تسمحان لنا بالوصول إلى هذه النتيجة : تقوم الأولى على أن نسحب الوحدات الأوّلية باحتمالات متساوية ، والثانية على أن نسحبها باحتمالات تتناسب مع أحجامها ، أي مع عدد الوحدات الثانوية التي تؤلّفها .

في كلتي الحالتين ، يجري تعيين الوحدات الثانوية داخل الوحدات الأوّلية باحتمالات متساوية .

1 . سحب الوحدات الأوّلية باحتمالات متساوية

تقوم هذه الطريقة :

- عند الدرجة الأولى ، على أن نسحب الوحدات الأوّلية بإعطائنا كلاًّ منها نفس احتمال التعيين p_1 ؛
- عند الدرجة الثانية ، على أن نسحب في كلّ وحدة عيّنة أوّلية ، الوحدات الثانوية بإعطائنا كلاًّ منها نفس احتمال الاختيار p_2 .

إذن ، كلّ وحدة ثانوية لها نفس الاحتمال p_1p_2 لأن تنتمي إلى العيّنة . ويساوي هذا الاحتمال معدّل أو نسبة البحث الإحصائي الأخيرة t :

$$t = p_1p_2$$

مثلاً . نريد أن نسحب على درجتين عيّنة من الأراضي الزراعية ، بمعدّل بحث إحصائي يساوي 1/100 ، وحيث تتكوّن الدرجة الأولى من البحث من عيّنة من المقاطعات .

يمكننا مثلاً أن نسحب باحتمالات متساوية مقاطعة على خمس ($p_1 = 1/5$) وفي كلّ مقاطعة ـ عيّنة ، أرضاً زراعية على عشرين ($p_2 = 1/120$) . معدّل البحث الإحصائي النهائي هو بالفعل :

$$t = \frac{1}{5} \times \frac{1}{20} = \frac{1}{100},$$

لكلّ أرض فرصة واحدة على مئة لأن تقع القرعة عليها .

نلاحظ أن عدد الوحدات الثانوية التي تنتمي الى العينة هو عدد عشوائي

ويساوي أمله الرياضي Nt ، حيث N تمثّل عدد الوحدات الثانوية الإجمالي ، ويكون تباينه أعلى كلّما كان عدد الوحدات الأوّلية أقل وأحجامها أكثر تفاوتاً . إنّ هذه الطريقة تعطي نتائج غير دقيقة عندما يكون حجم الوحدات الأوّلية كثير التغيّر . بالتالي من المستحسن أن يتمّ قبل السحب ، تجميع الوحدات الصغيرة وتقطيع الوحدات الكبرى بشكل نحصل فيه على وحدات أوّلية بأحجام متقاربة .

2 . سحب الوحدات الأوّلية باحتمالات تتناسب مع أحجامها

ـ عند الدرجة الأولى ، نسحب مع ردّ m وحدة أوّلية بإعطائنا كلاً منها احتمالاً لأن تُعيّن يتناسب مع عدد الوحدات الثانوية التي تؤلّفها N_a .

ـ عند الدرجة الثانية ، نسحب دون ردّ من كلّ وحدة عيّنة أوّلية العدد نفسه n_0 من الوحدات الثانوية .

مثلاً : تتضمّن إحدى المناطق ، المقسّمة إلى 20 مقاطعة ، 10400 أرض زراعية . نريد أن نسحب على درجتين عيّنة من الأراضي الزراعية بمعدّل بحث إحصائي يساوي 1/100 ، وحيث تتكوّن الدرجة الأولى من البحث من عيّنة من المقاطعات .

يمكننا مثلاً سحب 4 مقاطعات باحتمالات تتناسب مع عدد الأراضي الزراعية في كلّ منها .

العدد الإجمالي للأراضي المعيّنة يجب أن يكون : $10400 \times 1/100 = 104$.

إذن ، عند الدرجة الثانية نسحب عيّنة من (104/4=26) 26 أرضاً زراعية من كلّ مقاطعة معيّنة .

بطريقة السحب هذه يكون عدد الوحدات الثانوية n التي تنتمي إلى العيّنة عدداً ثابتاً . وهو يساوي :

$$n = m\, n_0$$

لكلّ الوحدات الثانوية نفس احتمال الانتماء إلى العيّنة ، وهذا الاحتمال يساوي معدّل البحث الإحصائي النهائي t :

$$t = \frac{mn_0}{N},$$

حيث N يمثّل عدد الوحدات الثانوية الإجمالي في المجتمع الإحصائي .

في الواقع ، بالنسبة لكلّ من السحوبات الـ m التي نجريها مع ردّ عند درجة البحث الإحصائي الأولى ، فإنّ احتمال ظهور الوحدة الأوّلية U_a هو N_a/N . بالنسبة لمجموعة السحوبات الـ m ، يساوي هذا الاحتمال (قاعدة الاحتمالات الكلّية) :

$$m\frac{N_a}{N}.$$

بعد تعيين الوحدة U_a ، احتمال ظهور الوحدة الثانوية $U_{a\beta}$ عند الدرجة الثانية من البحث يساوي $n_0 : N_a$

الاحتمال $p_{a\beta}$ لأن تنتمي الوحدة الثانوية $U_{a\beta}$ إلى العيّنة هو حاصل ضرب هذين الاحتمالين (قاعدة الاحتمالات المركّبة) :

$$p_{a\beta} = m\frac{N_a}{N}\cdot\frac{n_0}{N_a} = \frac{mn_0}{N}.$$

بما أنّ السحب قد تمّ عند الدرجة الأولى مع ردّ إلى الوعاء ، قد نختار إحدى الوحدات الأوّلية عدّة مرّات ، مثلاً k مرّة . عملياً ، لا نسحب في هذه الوحدة k عيّنة مستقلّة تتألّف من n_0 وحدة ثانوية ، ولكن عيّنة واحدة مستنفِدة تتضمّن kn_0 وحدة ، بشكل لا يمكن معه اختيار نفس الوحدة الثانوية أكثر من مرّة واحدة . ويبقى احتمال الوحدة الثانوية في الانتماء إلى العيّنة مساوياً لـ mn_0/N .

عندما تكون الوحدات الأوّلية متفاوتة الحجم والأهمّية فإنّ الطريقة التي تقوم على سحبها باحتمالات تتناسب مع أحجامها تسمح بالحصول على تقديرات أدقّ بكثير من السحب باحتمالات متساوية . وهي تفترض وجود معلومات إضافية : عدد الوحدات

الثانوية في كلّ وحدة أوّلية وذلك بالنسبة لكلّ وحدات المجتمع الإحصائي الأوّلية .

ولكن ، عملياً ، يكفي أن نعرف هذا العدد على وجه التقريب .

مثلاً . N_a هو عدد الأراضي الزراعية المجهول في كلّ مقاطعة عند البدء بالحملة .

لا نملك سوى تقريب له N'_a ، عدد هذه الأراضي عند التعداد الأخير .

نسحب عند الدرجة الأولى الوحدات الأوّلية باحتمالات تتناسب مع :

$$\frac{N'_a}{N'} \quad \text{حيث} \quad N' = \sum_a N'_a .$$

ونضع في كلّ وحدة أوّلية تعيّنها القرعة ، من أجل درجة السحب الثانية ، لائحة الأراضي الزراعية : عندئذٍ نحيط علماً بـ N_a .

عند الدرجة الثانية ، لا نسحب من وحدة العيّنة الأوّلية U_a ، n_0 وحدة ثانوية ، بل :

$$\frac{n_0}{N'_a} . N_a .$$

ضمن هذه الشروط ، فإنّ احتمال وحدة ثانوية بالانتماء إلى العيّنة يساوي mn_0/N' . إذن يمكن دوماً تعداد العيّنة كما في فرز الأصوات . بالمقابل التغى وجود نفس عدد الوحدات الثانوية في كلّ وحدة أوّلية .

القسم II

المناهج المعتمدة في تحسين دقّة الأبحاث الإحصائية العشوائية

1 . التفريع : A . المبدأ ؛ B . كيفية تحديد الفروع ؛ C . الخصائص ؛ D . توزيع العينة الأمثل بين الفروع : عيّنة نيمان ؛ E . ربح الدقّة العائد إلى التفريع . - 2 . التفريع البعدي وتقويم العيّنة : A . مبدأ التفريع البعدي ؛ B . اختيار معايير التفريع ؛ C . خصائص التفريع البعدي ؛ D . تحقيق التعداد عملياً ؛ E . تقويم العيّنة : « عدم الإجابات » .

من ضمن كلّ الطرق التي تسمح بتحسين دقّة التقديرات الناجمة عن بحث إحصائي عشوائي ، هناك اثنتان على أهمّية خاصّة . الأولى سابقة لسحب العيّنة ، وتقوم على تقسيم المجتمع الإحصائي إلى عدد معيّن من المجموعات المتجانسة وعلى توزيع العيّنة بين هذه المجموعات بغية تخفيض تقلّبات المعاينة : إنها طريقة التفريع . والثانية التي تأتي في طور تعميم النتائج ، تقوم على استعمال معلومات إحصائية إضافية : إنّها طريقة التفريع البعدي أو اللاحق .

I . التفريع

A . المبدأ

يقوم التفريع على أن نقطّع المجتمع الإحصائي موضع الدراسة إلى مجموعات متجانسة ، نسمّيها فروعاً ، وعلى أن نسحب بشكل مستقلّ عيّنة عشوائية من كلّ فرع .

دائماً يأتي تفريع العيّنة ، حتّى بشكل غير كامل ، في صالحنا : لا يمكن إلاّ أن نربح في الفعالية ، في الواقع ، حتّى ولو كانت وحدات المجتمع الإحصائي موزّعة بالصدفة بين الفروع، فإنّ العيّنة مأخوذة بواسطة سحب مستنفد مع معدّل بحث متماثل في كلّ فرع ، نفس دقّة عيّنة نموذجية بنفس الحجم . إذن ليس هناك من تضادّ .

B . كيفية تحديد الفروع

يقوم التفريع والبحث الإحصائي بالأنصبة أو بالكوتا على نفس الفكرة : وهي الحصول ، بواسطة فحص بعض المتغيّرات ، على عيّنة تكون صورة ، صادقة قدر الإمكان ، عن المجتمع الإحصائي . ويكمن الفارق ـ وهو جوهري ـ في كون الباحث هو من يختار العيّنة بالنسبة لطريقة الكوتا ، في حين أنّ القرعة هي من يختارها في كلّ فرع بالنسبة لطريقة البحث الإحصائي العشوائي المفرّع .

أ ـ اختيار معايير التفريع

يخضع اختيار معايير الفحص التي سنستخدمها لتحديد الفروع ، لاعتبارات شبيهة بالتي طرحناها بصدد طريقة الكوتا (الفصل V ، ص 222) .

كي يتسنّى اختيار خاصّة إحصائية ، كمّية أو نوعية ، كمعيار للتفريع ، يجب أن :

ـ تكون على ارتباط وثيق مع المتغيّرات موضع الدراسة . في الواقع ، تتعلّق فعالية

التفريع بتجانس الفروع إزاء هذه المتغيّرات . إذاً ، يتمّ اختيار معايير التفريع تبعاً للدراسة المشروع بها ؛

- يكون لها قيمة معروفة ، قبل الحملة ، بالنسبة لكل وحدات المجتمع الإحصائي إذاً لا يكفي ، كما بالنسبة لطريقة الكوتا ، أن نعرف توزيع هذه الخاصّة الإحصائي في مجمل المجتمع الإحصائي .

إذا لم نتوصّل إلى معرفة المعيار على وجه الدقّة ، فإنّ أخطاء التصنيف التي قد تُرتكب عند تكوين الفروع ، يمكنها ، بتنقيصها من تجانس هذه الفروع ، أن تخفّض من فعالية الطريقة ؛ ولا يحتمل أبداً أن تكون ، كما في طريقة الكوتا ، سبباً لتحيّز ما : إنّها ميزة حاسمة لطريقة الأبحاث الإحصائية العشوائية .

ويمكن استعمال التفريع عند كلّ درجة من بحث إحصائي بعدّة درجات (التفريع الثانوي) .

أمثلة

- حملة حول المقاصد الشرائية للأسر . بحث إحصائي بدرجتين .

أ - معايير تفريع الوحدات الأوّلية (المقاطعات أو التجمّعات السكنية) : المنطقة الجغرافية ، عدد السكّان ، نسبة السكّان الذين يعيشون من الزراعة ؛ .

ب - معايير التفريع الثانوي للوحدات الثانوية (الأسر) : الحيّ في المدن ، حجم الأسرة ،الفئة الاجتماعية - المهنية لربّ الأسرة .

- حملة دراسة صناعيّة . بحث إحصائي بدرجة واحدة .

معايير التفريع : حجم المؤسّسة (عدد الموظّفين ، مجموع المبيعات) ، فرع النشاط الاقتصادي .

ب - اختيار حدود وعدد الفروع

لقد كانت مسألة تجزئة المجتمع الإحصائي إلى فروع - اختيار حدود الفروع وعددها - موضوع أعمال نظرية ، خاصة أعمال دالينيوس (Dalenius) . وتؤدّي بنا الشروط التي وجدت إلى حسابات صعبة التطبيق ؛ ويتمّ بشكل عام حلّ هذه المسألة بطريقة تجريبية جدّاً انطلاقاً من بعض الأفكار الموجّهة البسيطة . بصورة خاصّة، لقد أظهرت بعض الدراسات النظرية والاختبارية أنّ مضاعفة عدد الفروع يأتي في صالحنا ، طالما تكون كلفة التفريع ضعيفة على العموم . ويجدر بالطبع أن نقف عند

ضرورة أن نسحب على الأقلّ وحدة ـ عيّنة من كلّ فرع ، وعلى الأقلّ وحدتين إذا كنّا نرغب في حساب دقّة التقدير .

في الحقيقة ، يتناقص مردود العملية بسرعة إذا ضاعفنا عدد الطبقات بالنسبة لكلّ معيار للتفريع : إذن ، نادراً ما نكوّن أكثر من سبعة أو ثمانية فروع انطلاقاً من نفس الخاصّة .

C . الخصائص

يسمح التفريع بتحسين دقّة التقديرات إلى حدّ بعيد ، بالنسبة لكلفة ضعيفة عادة ، طالما يكون من الممكن تحديد توزيع أمثل للعيّنة بين الفروع . ولكن ، بشكل عام ، لا يعود بإمكاننا تعداد النتائج كما بالنسبة لفرز الأصوات : يجب ترجيح كلّ مشاهدة بمعكوس معدّل البحث الإحصائي بالنسبة للفرع الذي تنتمي إليه .

فقط في الحالة حيث يكون معدّل البحث الإحصائي متماثلاً في كلّ فرع يمكننا إجراء التعداد كما لفرز الأصوات . ولكن عيّنة كهذه ، ونسميها عيّنة مفرّعة ممثلة ، لا تعطي بشكل عام سوى دقّة أضعف بكثير ولو أنّه لا يمكن إغفالها .

لنأخذ مجتمعاً إحصائياً مقداره N مقطّعاً إلى k فرع ، ونأخذ منه عيّنة بواسطة سحب مستنفِد . ولنأخذ متغيّرة X ننوي تقدير متوسّطها .

الرموز . سوف نعتمد الرموز التالية :

	الفروع							المجموعة
	1	2	...	h	...	k		
في المجتمع الإحصائي :								
المقدار	N_1	N_2	...	N_h	...	N_k		N
المتوسّط	m_1	m_2	...	m_h	...	m_k		m
التباين	σ_1^2	σ_2^2	...	σ_h^2	...	σ_k^2		σ^2
في العيّنة :								
المقدار	n_1	n_2	...	n_h	...	n_k		n
المتوسّط	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_h$	...	$\bar{x}_k$		$\bar{x}$
التباين	s_1^2	s_2^2	...	s_h^2	...	s_k^2		s^2

مع :

$$m_h = \frac{1}{N_h} \sum_{s=1}^{N_h} X_{hs}, \qquad \sigma_h^2 = \frac{1}{N_h} \sum_{s=1}^{N_h} (X_{hs} - m_h)^2$$

$$\bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}, \qquad s_h^2 = \frac{1}{n_h} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2 .$$

حيث :

X_{hs} تمثّل قيمة المتغيّرة X بالنسبة للوحدة الإحصائية U_{hs} ذات الرقم s داخل الفرع h ؛

x_{hi} تمثّل قيمة المتغيّرة X بالنسبة لوحدة العيّنة U_{hi} المعيّنة عند السحب رقم i في الفرع h .

أ ـ مقدّر متوسّط المجتمع الإحصائي

نقدّر متوسّط المجتمع الإحصائي :

$$m = \sum_{h=1}^{k} \frac{N_h}{N} m_h$$

بواسطة [1] :

$$\bar{x} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h = \frac{1}{N} \sum_{h=1}^{k} \frac{N_h}{n_h} \sum_{i=1}^{n_h} x_{hi} .$$

في الواقع ، بفضل خصائص الأمل الرياضي (الفصل 1 ، ص 55) :

$$E(\bar{x}) = E\left\{\sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h\right\} = \sum_{h=1}^{k} \frac{N_h}{N} E\{\bar{x}_h\} .$$

إلاّ أنّ الأمل الرياضي لمتوسّط عيّنة نموذجية يساوي متوسّط المجتمع الإحصائي الذي سُحبت منه (الفصل السادس ، ص 242) :

$$E\{\bar{x}_h\} = m_h .$$

(1) وليس بواسطة متوسّط العيّنة :

$$\bar{x} = \sum_{h=1}^{k} \frac{n_h}{n} \bar{x}_h .$$

المعاملات N_h/N تمثّل أوزان مختلف الفروع في المجتمع الإحصائي ، والمعاملات n_h/n أوزانها في العيّنة . في المعادلة $\bar{x}$ ، نرجّح كلّ مشاهدة x_{hi} بمعكوس معدّل البحث الإحصائي (أي بـ N_h/n_h) الخاصّ بالفرع الذي تنتمي إليه .

بالتالي :

$$E(\bar{x}) = \sum_{h=1}^{k} \frac{N_h}{N} m_h = m;$$

$\bar{x}'$ هو مقدّر غير متحيّز لمتوسّط المجتمع الإحصائي .

ب ـ العيّنة المفرّعة الممثّلة

من المستحسن عملياً الحصول على عيّنة مفرّعة يمكننا تعدادها كما فرز الأصوات ، دون أن يكون من الضروري إعطاء ترجيحات مختلفة لمختلف المشاهدات الفردية : عندها نقدر متوسّط المجتمع الإحصائي الكلّي بواسطة المتوسّط المناسب المحسوب على العيّنة ، ونقدّر النسبة بواسطة التردّد المناسب الملحوظ على العيّنة . كيف يجب توزيع العيّنة بين الفروع للوصول إلى هذه النتيجة ؟

بشكل عام ، في بحث إحصائي مفرّع ، نقدّر متوسّط المجتمع الإحصائي m بواسطة :

$$\bar{x}' = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h = \sum_{h=1}^{k} \frac{N_h}{N} \cdot \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} .$$

كي يمكن تعداد العيّنة كفرز الأصوات ، يجب أن يكون بوسعنا تقدير متوسّط المجتمع الإحصائي m بواسطة متوسّط العيّنة :

$$\bar{x} = \frac{1}{n} \sum_{h=1}^{k} \sum_{i=1}^{n_h} x_{hi} .$$

بالتالي ، الشرط الضروري والكافي كي يكون بوسعنا تعداد عيّنة مفرّعة كفرز الأصوات هو :

$$\frac{N_h}{N} \cdot \frac{1}{n_h} = \frac{1}{n}$$

أي :

$$\frac{n_h}{N_h} = \frac{n}{N} = t .$$

إذن يجب سحب العيّنة بمعدّل بحث إحصائي t متماثل بالنسبة لمختلف الفروع . ونسمّي عيّنة كهذه عيّنة مفرّعة ممثّلة .

ج - تباين مقدِّر المتوسِّط

بما أنَّ سحب العيِّنة يتمّ بشكل مستقلّ في كلّ فرع ، فإنّ مقدِّر المتوسِّط :

$$\bar{x} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h$$

هو حاصل جمع k متغيِّرة عشوائية مستقلّة . وبفضل خصائص التباين (الفصل I ، ص 60) :

$$V(\bar{x}) = V\left\{ \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h \right\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2} V(\bar{x}_h).$$

وبما أنَّ السحوبات في كل فرع تمّت بدون ردّ :

$$V(\bar{x}_h) = \frac{N_h - n_h}{N_h - 1} \cdot \frac{\sigma_h^2}{n_h}.$$

بالتالي ، تباين المقدِّر هو :

$$V(\bar{x}) = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h - 1} \cdot \frac{\sigma_h^2}{n_h}.$$

في هذه العبارة يمكننا تقدير σ_h^2 ، المجهولة ، وبدون تحيّز (أنظر الفصل VI ، ص 250) بواسطة :

$$\frac{N_h - 1}{N_h} s_h'^2$$

حيث :

$$s_h'^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2.$$

وإذا استبدلنا σ_h^2 بواسطة تقديرها ، نحصل على تقدير غير متحيّز لـ $V(\bar{x})$:

$$V^*(\bar{x}) = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h} \cdot \frac{s_h'^2}{n_h}.$$

د - تقدير النسبة

يمكننا مباشرة تعميم النتائج المتعلّقة بتقدير المتوسّط إلى تقدير نسبة خاصّة معيّنة في المجتمع الإحصائي ، باعتبارنا المتغيّرة X كمتغيّرة برنولي (أنظر الفصل VI ، ص 250) تأخذ القيمة 1 عندما تملك الوحدة الإحصائية موضع الدراسة هذه

الخاصّة ، والقيمة صفر عندما لا تملكها .

لنفترض :

$$p_1, p_2, \ldots, p_h, \ldots, p_k \; : \; p$$

نسبة الوحدات الإحصائية التي تملك الخاصّة موضع السؤال في كلّ من الفروع وفي مجمل المجتمع الإحصائي ؛

$$f_1, f_2, \ldots, f_h, \ldots, f_k \; : \; f$$

الترددات المناسبة الملحوظة على العيّنة .

بما أنّ X هي متغيّرة برنولي ، لدينا :

(متوسّط القيم X_{hi} في المجتمع الإحصائي)

$$p_h = \frac{1}{N_h} \sum_{i=1}^{N_h} X_{hi}$$

(متوسّط القيم x_{hi} في العيّنة)

$$f_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}$$

بالتالي ، نقدّر النسبة :

$$p = \sum_{h=1}^{k} \frac{N_h}{N} p_h$$

بواسطة :

$$f' = \sum_{h=1}^{k} \frac{N_h}{N} f_h .$$

تباين المقدّر هو :

$$V(f') = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h - 1} \cdot \frac{p_h(1 - p_h)}{n_h} .$$

وذلك بما أنّ تباين متغيّرة برنولي داخل الفرع h يساوي :

$$\sigma_h^2 = p_h(1 - p_h) .$$

ونقدّر تباين المقدّر بدوره بواسطة :

$$V^*(f') = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h} \cdot \frac{f_h(1 - f_h)}{n_h - 1}$$

D . توزيع العيّـنة الأمثل بين الفروع : عيّـنة نيمان Neyman

إذا كنّـا نرغب بتعداد العيّـنة كالفرز ، ينبغي سحبها بمعـدّل بحث متماثـل في مختلف الفروع .

ولكن يمكننا ، بالمقابل ، أن نسعى إلى التوزيع بين مختلف الفروع ، للعيّـنة ذات المقدار المثبّـت n ، بشكل نحصل فيه على أفضل تقدير ممكن .

هذه المسألة هي مسألة حدّ أدنى مرتبط : المقصود هو تحديد أحجام العيّـنات التي علينا سحبها من مختلف الفروع ، أي الأحجام n_1 ، n_2 ، . . . ، n_k التي تجعل تباين المقدّر

$$V\{\bar{x}\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h - n_h}{N_h - 1} \cdot \frac{\sigma_h^2}{n_h}$$

حدّاً أدنى ، مع الشرط :

$$\sum_{h=1}^{k} n_h = n .$$

وهذا يعني (أنظر تذكير الرياضيـات ، القسم I ، ص 318) أن نبحث عن الحدّ الأدنى للعبارة التالية :

$$W(n_1, \ldots, n_k) = V\{\bar{x}\} + \lambda\left(\sum_{h=1}^{k} n_h - 1\right)$$

حيث λ هو مضروب لاغرانج (Lagrange) .

نحصل على القيم n_h التي تناسب هذا الحدّ الأدنى بتصفيرنا الـ k مشتقّة جزئية التالية :

$$\frac{\partial W}{\partial n_h} = -\frac{N_h^2}{N^2} \cdot \frac{N_h}{N_h - 1} \cdot \frac{\sigma_h^2}{n_h^2} + \lambda = 0 , \qquad h = 1, 2, \ldots, k .$$

بشكل عام ، يكون حجم الفروع كبيراً بشكل يكفي لجعل :

$$\frac{N_h}{N_h - 1} \simeq 1$$

عندئذٍ يمكننا كتابة المعادلات السابقة :

$$-\frac{N_h^2}{N^2} \cdot \frac{\sigma_h^2}{n_h^2} + \lambda = 0 , \qquad h = 1, 2, \ldots, k$$

ما يعطي :

$$n_h^2 = \frac{1}{N^2 \lambda} N_h^2 \sigma_h^2 .$$

إذا وضعنا : $k = 1/N \sqrt{\lambda}$ ، نحصل على الشرط التالي :

$$n_h = K N_h \sigma_h . \qquad (1)$$

هذه العلاقة تعني أنّه يجب أن نختار في كلّ فرع عيّنة يكون مقدارها تناسبياً في آن واحد مع حجم الفرع وانحراف المتغيّرة موضع الدراسة X النموذجي في هذا الفرع .

لنرمز بواسطة :

$$t_h = \frac{n_h}{N_h}$$

إلى معدّل البحث الإحصائي في الفرع h . الشرط (1) يصبح :

$$t_h = K \sigma_h .$$

إذن للحصول على أفضل تقدير ممكن ، ينبغي أن يتناسب معدّل البحث الإحصائي في كلّ فرع مع الانحراف النموذجي للمتغيّرة موضع الدراسة في هذا الفرع .

بالتالي ، وكما يملي عليه الحدس ، يجب أن يكون معدّل البحث الإحصائي مرتفعاً أكثر كلّما كان تشتّت المتغيّرة موضع الدراسة داخل الفرع أكبر .

ونحدّد قيمة مُعَامِل التناسبية k بواسطة معادلة الارتباط :

$$\sum_{h=1}^{k} n_h = K \sum_{h=1}^{k} N_h \sigma_h = n$$

إذن :

$$K = \frac{n}{\sum_{h=1}^{k} N_h \sigma_h} .$$

ونسمّي العيّنة التي نختارها بهذه الطريقة عيّنة نيمان (Neyman) نسبة إلى اسم مبتكر هذه الطريقة .

وضع الطريقة موضع التنفيذ

إنّ التوزيع الأمثـل للعيّـنة بـين الفروع يفتـرض أنّنا نعـرف انحرافـات المتغيّـرة موضع الدراسة النموذجية في كلّ فرع . في الحقيقة لا نملك بشكل عام أكثر من فكـرة تقريبية عن هذه الانحرافات .

من جهـة أخرى ، لا تقتصر الـدراسة عـادةً على متغيّـرة واحـدة . والعيّـنة التي تكون مثلى بالنسبة لتقدير متوسّـط X ، قد لا تكون كذلك بالنسبة لِـ Y .

أكثر الأحيان ، نحلّ هذه المسائل بتحديدنا الفروع من خلال « حجم » الوحدات وكذلك بتحديدنا التوزيـع الأمثل للعيّـنـة بالنـسـبة لتقديـر متوسّـط الأحجـام . وبما أنّ المتغيّـرات الكمّية المعرّضة للدراسة هي بشكل عام على ارتباط وثيق مـع « الحجم » ، تصبح العيّـنة التي نضعها بهذه الطريقة جيّـدة أيضـاً بالنسـبة لتقديـر متوسّـطات هـذه المتغيّـرات .

وكثيراً ما نستنتـج أن متوسّـطات متغيّـرة كمّية معيّـنة وانحرافـاتها النمـوذجية المتعلّـقة بمختلف الفروع هي تناسبية :

$$\frac{\sigma_h}{\overline{X}_h} = \text{ثابتة}$$

عندئذٍ ، تصبح قاعدة توزيع العيّـنة بين الفروع :

$$n_h = K' N_h \overline{X}_h$$

حيث $N_h \overline{X}_h$ يمثّـل حاصل المتغيّـرة X في الفرع h . من هنا القاعدة التجريبية التي تُستعمل كثيراً : تتوزّع العيّـنة بـين الفروع تنـاسبيـاً مـع مجموع المتغيّـرة المستعملة للتفريع .

مثلاً . ننوي القيام بحملة حول عيّـنة تتكوّن من 1000 مُؤسّـسة صناعية للحصول على معلومات عن الانتـاج ، القيمة المضـافة والاستثمـارات . يتمّ تقـطيـع المجتمـع الإحصائي إلى فرعين ، فرع المؤسّـسات الكبيرة وفرع المؤسّـسات الصغيرة . نعرف على وجه التقريب مجموع عدد الموظّـفين في كلّ فرع . بما أنّ المتغيّـرات موضع الدراسة هي على ارتباط وثيق بعدد الموظّـفين ، فإنّنا نوزّع العيّـنـة تناسبيـاً مع هـذا العدد في كـلّ فرع :

	تحديد الفرع	عدد المؤسّسات في الفرع N_h	مجموع عدد الموظّفين في الفرع	مقدار العيّنة n_h
الفرع 1	مؤسّسات بـ50 موظّفاً وأكثر	2000	500000	625
الفرع 2	مؤسّسات بأقلّ من 50 موظّفاً	25000	300000	375
حواصل الجمع		27000	800000	1000

E . ربح الدقّة العائد إلى التفريع

لنأخذ عيّنة مفرّعة تبعاً لخاصّة A ، كمّية أو نوعية . ننوي تقدير m وهو متوسّط المتغيّرة X في المجتمع الإحصائي .

مقدِّر m هو :

$$\bar{x} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h$$

وفي حال عدم التفريع نقدِّر m بواسطة $\bar{x}$ ، وهو متوسّط X في العيّنة غير المفرّعة .

المقصود هو إذن مقارنة تبايني المقدِّرين $\bar{x}$ و $\bar{x}$ اللذين يناسبان عيّنتين بنفس الحجم . الربح العائد إلى التفريع هو :

$$G = V\{\bar{x}\} - V\{\bar{x}\}.$$

تباين المقدِّر غير المفرَّع

إذا افترضنا سحب العيّنة قد تمّ دون ردّ ، فإنّ تباين المقدِّر في حالة عيّنة غير مفرّعة ، هو (الفصل VI ، ص 247) :

$$V\{\bar{x}\} = \frac{N-n}{N-1} \cdot \frac{\sigma^2}{n}. \qquad (1)$$

ولكن إذا أدخلنا التقطيع إلى فروع ، يمكننا تجزئة تباين X في المجتمع الإحصائي

إلى حاصل جمع عنصرين ، تباين متوسّطات الفروع (التباين بين الفروع) ومتوسّط تباينات الفرع (التباين داخل الفروع) (راجع الكتاب الأوّل : « الإحصاء الوصفي » ، الفصل VI ، القسم II ، الفقرة 4.C) :

$$\sigma^2 = \sum_{h=1}^{k} \frac{N_h}{N}(m_h - m)^2 + \sum_{h=1}^{k} \frac{N_h}{N}\sigma_h^2 .$$

بالتالي ، يمكننا كتابة تباين المقدّر :

$$V\{\bar{x}\} = \frac{N-n}{N-1}\cdot\frac{1}{n}\left[\sum_{h=1}^{k} \frac{N_h}{N}(m_h - m)^2 + \sum_{h=1}^{k} \frac{N_h}{N}\sigma_h^2\right] \qquad (2)$$

تباين المقدّر المفرّع

تباين المقدّر ، في حالة عيّنة مفرّعة ، هو :

$$V\{\bar{x}'\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2}\cdot\frac{N_h - n_h}{N_h - 1}\cdot\frac{\sigma_h^2}{n_h} . \qquad (3)$$

إذا كانت معدّلات البحث الإحصائي في مختلف الفروع متساوية (العيّنة المفرّعة الممثّلة) :

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_k}{N_k} = \frac{n}{N} = t ,$$

تصبح العبارة (3) إذا استبدلنا n_h/N_h بواسطة n/N :

$$\begin{aligned} V\{\bar{x}'\} &= \sum_{h=1}^{k} \frac{N_h}{N^2}\cdot\frac{N_h - n_h}{N_h}\cdot\frac{N_h}{N_h - 1}\cdot\frac{N_h}{n_h}\sigma_h^2 \\ &= \sum_{h=1}^{k} \frac{N_h}{N^2}\cdot\frac{N - n}{N}\cdot\frac{N_h}{N_h - 1}\cdot\frac{N}{n}\sigma_h^2 \\ &= \frac{N - n}{N}\cdot\frac{1}{n}\sum_{h=1}^{k} \frac{N_h}{N}\cdot\frac{N_h}{N_h - 1}\sigma_h^2 \end{aligned}$$

إذن في حالة عيّنة مفرّعة ممثّلة ، تباين المقدّر هو :

$$V\{\bar{x}'\} = \frac{1 - t}{n}\sum_{h=1}^{k} \frac{N_h}{N}\cdot\frac{N_h}{N_h - 1}\sigma_h^2 . \qquad (4)$$

الربح العائد إلى التفريع

بشكل عام ، من غير الممكن اختزال عبارة الربح العائد إلى التفريع :

$$G = V\{\bar{x}\} - V\{\bar{x}'\}$$

لكن هذه العبارة تأخذ ، في حالة بحث إحصائي مفرّع ممثّل وعلى أساس بعض التقريبات ، شكلاً بسيطاً وإيحائياً خاصّاً .

في الواقع ، إذا كانت القيم N وN_h كبيرة ، يمكننا استبدال $1/(N-1)$ بـ $1/N$ و$1/(N_h-1)$ بـ $1/N_h$ في العبارتين (2) و(4) وكتابة :

$$V\{\bar{x}\} \approx \frac{1-f}{n}\left[\sum_{h=1}^{k}\frac{N_h}{N}(m_h-m)^2+\sum_{h=1}^{k}\frac{N_h}{N}\sigma_h^2\right]$$

$$V\{\bar{x}'\} \approx \frac{1-f}{n}\sum_{h=1}^{k}\frac{N_h}{N}\sigma_h^2 .$$

إذن ، الربح العائد إلى التفريع هو في هذه الحالة :

$$G = V\{\bar{x}\} - V\{\bar{x}'\} \approx \frac{1-f}{n}\sum_{h=1}^{k}\frac{N_h}{N}(m_h-m)^2$$

والربح النسبي :

$$\frac{V\{\bar{x}\} - V\{\bar{x}'\}}{V\{\bar{x}\}} \approx \frac{\sum_{h=1}^{k}\frac{N_h}{N}(m_h-m)^2}{\sigma^2} = \eta^2_{X/A}$$

يساوي (أنظر الفصل IV ، ص 177) نسبة ارتباط X بـ A ، حيث A هي الخاصّة المعتمدة كمعيار للتفريع .

وبما أنّ الربح العائد إلى التفريع يساوي صفراً إذا كان $\eta^2_{X/A}=0$ ، فهو على أهمّية أكبر كلّما كان ارتباط X مع A وثيقاً أكثر . وعندما يكون $\eta^2_{X/A}=1$ ، يكون التقدير دقيقاً تماماً ، لأنّ التباين داخل كلّ فرع يساوي عندئذٍ صفراً .

بالمختصر :

- من صالحنا دائماً أن نفرّع . حتّى ولو لم يكن بإمكاننا تحديد التوزيع الأمثل للعيّنة بسبب جهلنا للانحرافات النموذجية ، داخل كلّ فرع ، للمتغيّرة المستعملة كمعيار للتفريع ، إذ أنّ تفريعاً بمعدّل بحث إحصائي متماثل («العيّنة المفرّعة الممثّلة») هو أفضل من عدم التفريع
- يكون الربح العائد إلى التفريع أقوى كلّما كان ارتباط المتغيّرة موضع الدراسة مع معيار التفريع وثيقاً أكثر ، بعبارة أخرى كلّما كانت الفروع ، من وجهة نظر المتغيّرة موضع الدراسة ، مختلفة أكثر عن بعضها البعض .

الجدول 25 . مقارنة فعالية مختلف طرق التفريع

الفرع	مقدار الفرع N_h	مقدار العيّنة n_h			
		لا تفريع	العيّنة المفرّعة المثلى	العيّنة المثلى	التوزيع التناسبي مع مجموع المبيعات
1	538		15	244	286
2	4756		131	288	409
3	30964		854	468	305
المجموع	36258	1000	1000	1000	1000
معامل تغيّر المقدّر	$\frac{\sigma(\bar{x})}{m}$	9,9%	7,1%	3,0%	3,3%

لإعطاء فكرة عن الربح العائد إلى التفريع ، نجد أعلاه (الجدول 25) نتائج إحدى دراسات هانسن Hansen وهورفيتز Hurwitz ، ذكرها J. Desabie [1] . وتتعلّق هذه النتائج ببحث إحصائي جرى حول مؤسّسات صناعية مفرّعة حسب مجموع مبيعات السنة المنصرمة . ونجري المقارنة في ما يخصّ معامل تغيّر متوسّط الراتب الموزّع .

نلاحظ أهمّية الربح العائد إلى إستعمال عيّنة مثلى . كما نلفت إلى أنّ الطريقة التجريبية في توزيع العيّنة الأمثل يؤدّي أيضاً إلى نتائج مرضية كثيراً .

2 . التفريع البعدي وتقويم العيّنة

A . مبدأ التفريع البعدي

يقوم التفريع البَعدي أو اللاحق على تحديد الفروع بعد سحب العيّنة وعلى ترجيح ، كما في التفريع السابق ، كلّ من المشاهدات بواسطة مُعامل تناسبي مع مقدار الفرع في المجتمع الإحصائي .

إذن يستدعي التفريع البعدي الإحاطة بفكرة إضافية : وهي توزيع المجتمع

(1) J. Desabie, Théorie et pratique des sondages. Dunod , 1971

الإحصائي بين الفروع . وهذه الضرورة هي أضعف بكثير من الضرورة التي يفرضها التفريع السابق حيث يستدعي معرفة قيمة ، أو كيفية ، معيار التفريع بالنسبة لكلّ وحدات المجتمع الإحصائي .

الدقّة التي نحصل عليها بواسطة التفريع البعدي هي أكبر من دقّة عيّنة غير مفرّعة . ولكنّها ، بالمتوسّط ، أصغر من دقّة عيّنة مفرّعة قبل السحب ، وحسب نفس تقطيع الفروع ، بمعدّل بحث إحصائي متماثل .

بالتالي ، من الأفضل دوماً اعتماد تفريع العيّنة قبل السحب . ونستعمل التفريع البعدي :

- عندما لا نحيط علماً بخاصّة التفريع بالنسبة لكلّ وحدات المجتمع الإحصائي ، فلا يسمح لنا بإجراء التفريع قبل السحب ؛
- عندما لا تظهر أهمّية التفريع حسب معيار معيّن إلاّ أثناء التشغيل ، بعد أن نكون قد استنتجنا ، مثلاً ، ارتباطاً قويّاً بين هذا المعيار والمتغيّرة موضع الدراسة .

B . اختيار معايير التفريع

يخضع اختيار أحد معايير التفريع البعدي لنفس شروط اختيار متغيّرة مراقبة في بحث إحصائي بالأنصبة أو بالكوتا . يجب على معيار التفريع :

- أن يكون على ارتباط وثيق مع المتغيّرات موضع الدراسة ؛
- أن يكون توزيعه الإحصائي معروفاً في مجمل المجتمع الإحصائي ؛
- أن يكون قد تمّت مشاهدته أثناء الحملة دون إمكانية خطأ كبيرة .

وهذا الشرط الأخير هو على أهمّية ، فمن الضروري في الواقع أن نقوم بتصنيف وحدة إحصائية معيّنة في أحد الفروع حسب نفس القواعد المعتمدة لوضع الإحصائية التي نستخدمها لتحديد مقدار كلّ فرع . وإذا لم يكن الأمر كذلك يتأثّر تقدير كمّية معيّنة انطلاقاً من العيّنة ، كما في حالة البحث الإحصائي بالكوتا، بخطأ منهجي . كي نتجنّب هذه المخاطرة ، نصنّف غالباً وحدات العيّنة حسب قيمة معيار التفريع التي تظهر في قاعدة البحث الإحصائي نفسها .

مثلاً . من أجل الدراسات حول الأسر ، غالباً ما تتكوّن قاعدة البحث الإحصائي من سجلّ شهادات السكن الموضوعة أثناء الفرز السكاني الأخير . لنفترض أنّنا أخذنا كمعايير للتفريع البعدي الجنس ، فئة ربّ الأسرة الاجتماعية المهنية وعدد

أعضاء الأسرة : نحدّد الفروع بتلاقي هذه الميزات الثلاث . نجد مقدار كلّ فرع انطلاقاً من نتائج الفرز ونأخذ من قاعدة البحث الإحصائي ، بالنسبة لكلّ مسكن - عيّنة ، قيمة هذه الميزات الثلاث لحظة الفرز . بهذه الطريقة نتجنّب تباعدات التصنيف بين الفرز والحملة ، سواء عادت هذه الأخطاء الى أخطاء معيّنة أو إلى تغييرات حقيقية جرت بين هاتين العمليتين .

بشكل طبيعي ، تتناقص فعالية التفريع الموضوع بهذه الطريقة كلّما ابتعدنا عن تاريخ الفرز السكّاني ، لأنّ الارتباط بين قيمة المتغيّرة موضع الدراسة ، لحظة الحملة ، وقيمة معيار التفريع لحظة الفرز ، يمضي وهو يضعف .

من جهة أخرى ، لا يفيدنا إدخال تفريع بعدي حسب معيار معيّن A إلاّ إذا كان توزيع A في العيّنة متحرّفاً حقّاً بالتقلّبات العشوائية .

C . خصائص التفريع البعدي

لنفترض :

- X ، متغيّرة إحصائية ننوي تقدير متوسّطها m انطلاقاً من العيّنة ؛
- A ، خاصّة نوعية أو كمّية ، نعرف توزيعها في المجتمع الإحصائي .

يقوم التفريع البعدي ، بعد سحب العيّنة ، على تحديد الفروع انطلاقاً من الخاصّة A وعلى تجميع وحدات العيّنة حسب هذه الفروع .

أ - مقدّر متوسّط المجتمع الإحصائي

كما في حالة التفريع قبل السحب ،

$$\bar{x} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h$$

هو مقدّر غير متحيّز لمتوسّط المجتمع الإحصائي m .

وتشكّل الأوزان النسبية N_h/N المعلومات الإحصائية الإضافية الضرورية لتحسين دقّة التقدير .

إلاّ أنّه إذا كانت التخمينات N'_h التي بحوزتنا بالنسبة لمقادير العيّنات خاطئة (معلومات إحصائية غير صحيحة ، قديمة جدّاً ، أو تستعمل تحديدات غير التي استعملت لتوزيع وحدات العيّنة بين الفروع) ، فإنّ مقدّر المتوسّط :

$$\bar{x}' = \sum_{h=1}^{k} \frac{N'_h}{N'} \bar{x}_h$$

هو نفسه متحيّز .

N_h هو المقدار الحقيقي للفرع h ، يمكننا في الواقع كتابة :

$$\bar{x}' = \sum_{h=1}^{k} \frac{N_h}{N}\bar{x}_h + \sum_{h=1}^{k}\left(\frac{N'_h}{N'} - \frac{N_h}{N}\right)\bar{x}_h .$$

بالتالي ، بفضل خصائص الأمل الرياضي (الفصل I ، ص 55) :

$$E\{\bar{x}'\} = E\{\bar{x}\} + \sum_{h=1}^{k}\left(\frac{N'_h}{N'} - \frac{N_h}{N}\right)E\{\bar{x}_h\}$$

إذن :

$$E\{\bar{x}'\} = m + \sum_{h=1}^{k}\left(\frac{N'_h}{N'} - \frac{N_h}{N}\right)m_h .$$

العنصر الثاني يمثّل الخطأ المنهجي الذي يتأثّر به التقدير .

ب ـ تباين مقدّر المتوسّط

بما أنّ التقطيع إلى فروع يأتي بعد سحب العيّنة ، لا يمكن تحديد توزيع هذه العيّنة بين الفروع مسبقاً : عدد وحدات العيّنة nh في كلّ فرع هو متغيّرة عشوائية . من غير الممكن مثلاً أن نبحث عن توزيع العيّنة الأمثل بين الفروع .

إلاّ أنّه ما أن تُسحب العيّنة حتى تصبح القيم nh أعداداً ثابتة . عندها تكون العيّنة شبيهة بالضبط بعيّنة سابقة التفريع لها نفس التوزيع بين الفروع . إذن تباين مقدّر المتوسّط هو (راجع الفقرة 1.C ، ص 331) :

$$V\{\bar{x}/n_1, n_2, \ldots, n_k\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2}\cdot\frac{N_h - n_h}{N_h - 1}\cdot\frac{\sigma_h^2}{n_h} .$$

ولكن حتّى قبل سحب العيّنة ، يحقّ لنا أن نتساءل عن مدى دقّة المقدّر $\bar{x}$. لا يمكن إيجاد سوى دقّة متوسّطة (بالتحديد أمل تباين المقدّر الرياضي) لأنّ توزيع العيّنة بين الفروع ليس أكيداً ، بل عشوائياً ، يمكننا كتابة :

$$V\{\bar{x}\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2}\,\frac{N_h - n_h}{N_h - 1}\cdot\frac{\sigma_h^2}{n_h}$$

على النحو الآتي :

$$V\{\bar{x}\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2}\cdot\frac{N_h}{N_h - 1}\cdot\frac{\sigma_h^2}{n_h} - \sum_{h=1}^{k} \frac{N_h^2}{N^2}\cdot\frac{N_h}{N_h - 1}\cdot\frac{\sigma_h^2}{N_h} .$$

وهي دالّة خطّية تبعاً للكمّيات العشوائية $1/n_h$.

وإذا أخذنا أمل هذه العبارة الرياضي :

$$E\{V(\bar{X}')\} = \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h}{N_h - 1} \sigma_h^2 E\left\{\frac{1}{n_h}\right\} - \sum_{h=1}^{k} \frac{N_h^2}{N^2} \cdot \frac{N_h}{N_h - 1} \cdot \frac{\sigma_h^2}{N_h} .$$

ولكن تحت الشرط ،

$$n_h \neq 0 \quad \forall h \qquad (\text{مهما يكن } h)$$

وهو شرط يتحقّق أثناء تقطيع الفروع بعدياً ، يمكننا إثبات أنّ [1] :

$$E\left\{\frac{1}{n_h}\right\} \approx \frac{N}{N_h} \cdot \frac{1}{n} + \frac{N}{N_h}\left(\frac{N - N_h}{N_h}\right)\frac{1}{n^2} + \cdots$$

حيث العناصر المهمَلة هي كمّيات لا متناهية الصغر بدرجة أكبر .

أخيراً :

$$E\{V(\bar{X}')\} \approx \frac{N - n}{N} \cdot \frac{1}{n} \sum_{h=1}^{k} \frac{N_h}{N} \cdot \frac{N_h}{N_h - 1} \sigma_h^2 + \frac{1}{n^2} \sum_{h=1}^{k} \frac{N - N_h}{N} \cdot \frac{N_h}{N_h - 1} \sigma_h^2 .$$

العنصر الأوّل يساوي تباين العيّنة المفرّعة الممثّلة ، العنصر الثاني هو كمية لا متناهية الصغر حسب $1/n^2$ ، دائماً إيجابية .

إذن بالمتوسّط يكون تباين العيّنة المفرّعة بعدياً أكبر من تباين العيّنة المفرّعة الممثّلة . ويصبح الفارق ضعيفاً جدّاً ما أن يكون مقدار العيّنة كبيراً بشكل كاف .

D . تحقيق التعداد عملياً

بما أنّ توزيع العيّنة بين الفروع هو عشوائي ، يُستبعد أن يكون بإمكاننا تعداده كالفرز . ينبغي إذن أن نرجّح فعلاً متوسّط كلّ فرع $\bar{x}_h$ بواسطة المعامل N_h/N المناسب .

تقوم طريقة أولى على إعطاء كلّ مشاهدة معامل الترجيح التابع للفرع الذي تنتمي إليه . هذه الطريقة ، التي كان يصعب استعمالها عندما كانت التعدادات تجري على عتاد كتابي آلي ، أصبحت تُستعمل بكثرة اليوم بفضل الحاسب الإلكتروني .

---

(1) انظر ، مثلاً : J. Desabie, Théorie et pratique des sondages. Dunod, 1971 الفصل 8 ، ص 180.

وتقوم طريقة ثانية على مضاعفة بعض المشاهدات وحذف أخرى . لنفترض أنّ العيّـنة تحتوي n_h مشاهدة في الفرع h الذي يجب أن يحتوي $n'_h = t.N_h$ مشاهدة .

إذا كانت $n_h < n'_h$ ، نعيّـن بالصدفة $V_h = n'_h - n_h$ مشاهدة نضاعفها .

إذا كانت $n_h > n'_h$ نسحب بالقرعة $V_h = n_h - n'_h$ مشاهدة نحذفها .

هذه الطريقة تسهّـل الحسابات إلى حدّ بعيد : بعد المضاعفات والإلغاءات يمكن تعداد العيّـنة المقوّمة كما فرز الأصوات .

إلاّ أنّـه إذا كانت هذه الطريقة تعطي تقديرات غير متحيّـزة . فإنّ هذه التقديرات هي أقلّ دقّـة بعض الشيء من التقديرات الناتجة عن الطريقة الأولى : حتّى عند عدم إلغاء بعض المعلومات ، فإنّ السحب بالقرعة للمشاهدات التي يجب مضاعفتها يُدخل عامل تغيّـر إضافي .

مثلاً . لنفترض أنّنا أجرينا حملة حول الإستهلاك على عيّـنة عشوائيـة تتكوّن من 1000 أسرة . يسمح لنا الجدول 26 الموضوع بعد الحملة بمقارنة توزيع الأسر حسب فئة ربّ الأسرة الاجتماعية ـ المهنية ، في العيّـنة وفي مجمل المجتمع .

الجدول 26 . مقارنة توزيع الأسر في العيّـنة وفي المجتمع الإحصائي

فئة ربّ الأسرة الاجتماعية ـ المهنية	بنية العيّـنة (%)	بنية المجتمع الإحصائي (%)
1 . مزارعون وأجراء زراعيّون	9,0	9,9
2 . أرباب عمل صناعيون وتجاريون	8,8	8,1
3 . مهن حرّة ، كوادر عليا	5,3	5,1
4 . كوادر متوسّـطة	6,9	7,4
5 . موظّـفون	6,9	7,5
6 . عمّال	25,8	28,0
7 . نشاطات أخرى	4,1	4,4
8 . غير عاملين	33,2	29,6
المجموع	100,0	100,0

الطريقة الأولى

يقدّر متوسّط المجتمع الإحصائي m بواسطة :

$$\bar{x} = \sum_{h=1}^{k} \frac{N_h}{N} \bar{x}_h = \sum_{h=1}^{k} \frac{N_h}{N} \cdot \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} .$$

نرمز بواسطة :

$t = n/N$ إلى معدّل أو نسبة البحث الإحصائي ،

$n'_h = tN_h$ المقدار النظري الذي كان يمكن الحصول عليه لو تمّ تفريع البحث الإحصائي مسبقاً ،

لدينا :

$$\frac{N_h}{N} = \frac{n'_h}{n} .$$

بالتالي ، يمكننا كتابة مقدّر المتوسّط أيضاً على الشكل التالي :

$$\bar{x} = \frac{1}{n} \sum_{h=1}^{k} \frac{n'_h}{n_h} \sum_{i=1}^{n_h} x_{hi} .$$

هكذا ، نعيد المقدار الحقيقي n_h لكلّ فرع إلى المقدار النظري n'_h بضربنا كلّ مشاهدة بعامل الترجيح n'_h / n_h (الجدول 27) .

الطريقة الثانية

بما أنّ العيّنة تحتوي على 90 أسرة من المزارعين والأجراء الزراعيين بدلاً من 99 ، نختار منها 9 بالصدفة ونضاعفها .

كذلك ، بعد أن نجد في العيّنة 7 استجوابات زائدة تتعلّق بأرباب العمل الصناعيين والتجاريين ، نحذف منها 7 نسحبها بالصدفة . ويعطينا الجدول 28 عدد الاستجوابات التي يجب أن نضاعفها أو نلغيها في كلّ فئة .

الجدول 27 . حساب معاملات الترجيح .

فئة ربّ الأسرة الاجتماعية - المهنية	مقدار العيّنة n_h	المقدار النظري n'_h	معامل الترجيح n'_h/n_h
1 . مزارعون وأجراء زراعيون	90	99	1,10
2 . أرباب عمل صناعيون وتجاريون	88	81	0,92
3 . مهن حرّة ، كوادر عليا	53	51	0,96
4 . كوادر متوسّطة	69	74	1,07
5 . موظّفون	69	75	1,09
6 . عمّال	258	280	1,09
7 . نشاطات أخرى	41	44	1,07
8 . غير عاملين	332	296	0.89

الجدول 28 . تقويم العيّنة بواسطة مضاعفة وحذف الاستجوابات

فئة ربّ الأسرة الاجتماعية - المهنية	مقدار العيّنة	المقدار النظري n'_h	عدد الاستجوابات للمضاعفة r_h	عدد الاستجوابات للحذف $-r_h$
1 . مزارعون وأجراء زراعيون	90	99	9	
2 - أرباب عمل صناعيون وتجاريون	88	81		− 7
3 . مهن حرّة ، كوادر عليا	53	51		− 2
4 . كوادر متوسّطة	69	74	5	
5 . موظّفون	69	74	5	
6 . عمّال	258	280	22	
7 . نشاطات أخرى	41	44	3	
8 . غير عاملين	322	296		− 36
	1000	1000	45	− 45

E . تقويم العيّنة : « عدم الإجابات »

حتّى الآن ، لم نأخذ بعين الاعتبار سوى انحرافات العيّنة العائدة إلى التقلّبات العشوائية . ولكن يوجد ، في الحملات التي تقام بين الجمهور ، أسباب

تحريف مهمّة أخرى : إنّها « عدم الإجابات » . لم يكن مثلاً بالإمكان استجواب أشخاص غائبين عن منازلهم طيلة فترة الحملة أو تعذّر الاتّصال بهم ، والبعض الآخر رفض الإجابة .

بحكم هذه الإخفاقات ، نجد عيّنة « الإجابات » قد اختلفت عن العيّنة النظرية التي اختارتها الصدفة وقد تأثّر بنيتها بهذا الأمر لعدم وجود أي سبب لقبول الاستقلالية بين فعل الإجابة والمتغيّرات موضع الدراسة .

هكذا ، قد لا نفي حقّ الأسر التي تتألّف من فرد واحد في التمثيل ، لصعوبة الاتصال بهذا الفرد . قد يوجد أيضاً رابط واضح بين موضع الحملة والميل إلى الإجابة : مثلاً قد تصطدم حملة حول « العمل غير الرسمي » بالكثير من الرفض للإجابة عند الأشخاص الذين يمارسون هذا النوع من العمل .

ضمن هذه الشروط ، لا يعود بالإمكان اعتبار عيّنة « الإجابات » عيّنة عشوائية مسحوبة من مجمل المجتمع الإحصائي ويخشى عندها وجود تحيّز معيّن .

لنرمز في الواقع بواسطة p_s إلى احتمال ظهور إجابة الفرد U_s في العيّنة وبواسطة X_s إلى قيمة المتغيّرة موضع الدراسة .

بفضل خصائص الأمل الرياضي :

$$E(\bar{x}) = \sum_{s=1}^{N} p_s X_s .$$

عبارة التحيّز هي :

$$E(\bar{x}) - m = \sum_{s=1}^{N} \left(p_s - \frac{1}{N}\right) X_s = \sum_{s=1}^{N} \left(p_s - \frac{1}{N}\right)(X_s - m) .$$

إذن ، يساوي هذا التحيّز صفراً :

- إذا كانت احتمالات مختلف الوحدات للانتماء إلى العيّنة متساوية :

$p_s = \frac{1}{N}$ مهما تكن s ؛

- إذا كانت الاحتمالات p_s وقيم المتغيرة موضع الدراسة X_s مستقلّة .

لكن بشكل عام ، لا يتحقّق أيّ من هذين الشرطين :

- الاحتمالات p_s هي غير متساوية ومن جهة أخرى مجهولة ، وقد يكون البعض منها صفراً ؛

ـ غالباً ما يوجد ارتباط بين X_a وp_a .

بشكل عام ، يُظهر البحث النظري أنّـه في صالحنا أن نكرّس أقصى جهـدنا كي نحصل على إجابات كلّ وحدات العيّـنـة تقريبـاً ، مع احتمـال تخفيض مقدار العيّـنـة الأصلية كي نبقى في حدود ميزانية الحملة .

نستعمل عادة ثلاث طرق لتقويم عيّـنة « عدم الإجابات » ، وهي :

ـ التفريع البعدي ؛
ـ استبدال الأفراد المتخلّفين ؛
ـ استعمال عيّـنة ثانوية من غير المجيبين .

أ ـ تقويم العيّـنة بواسطة التفريع البعدي

يسمح التفريع البعدي بتصحيح بنية العيّـنة من التحريفات المنهجية العـائدة إلى « عدم الإجابات » كما من التحريفات العائدة إلى التقلّـبات العشوائية .

في صالحنا أن نأخذ كمعيار للتفريع متغيّـرة تكون في آن واحد :
ـ ذات توزيع حُرِّف بشكل واضح بحكم « عدم الإجابات » ،
ـ على ارتباط وثيق مع المتغيّـرات موضع الدراسة .

عادة ، هذه هي مثلاً حالة عدد أفراد الأسرة . ويمكننا طبعاً تبنّي عـدّة متغيّـرات مراقبة في نفس الوقت .

ب ـ استبدال الأفراد المتخلّفين

في بعض الحالات ، يكون توزيع المجتمع الإحصائي بين الفروع مجهـولاً أو غير معـروف على وجـه الصحّـة : مصـدر قـديم ، تحـديـدات مختلفـة عن التي تـستعملهـا الحملة ، أخطاء في المشاهدة ، الخ . . إذن لا يمكننا تقييد بنية العيّـنة بهذا التوزيـع . بالمقابل ، يمكننا « استبدال » كلّ فرد متخلّـف .

لهذا نختار ، كما بالنسبة للتفريـع البعدي ، متغيّـرة أو أكـثر للمراقبـة ونـسعى ، باستجوابنا الجيران مثلاً ، لتحديد قيمها بالنسبة لكلّ فرد متخلّـف . يمكننا عندئذٍ :

ـ إمّـا استبدال كلّ فرد متخلّـف بشخص له نفس الميزات ، نضاعف إجابته ؛
ـ إمّـا أن نلحق بأجوبة الأفراد اللذين يمثّـلون نفس ميزات المراقبة مُعامِل الترجيح الذي يعوّض عن الأفراد المتخلّـفين .

هذه الطريقة ، بعكس طريقة التفريـع البعدي ، لا تصحّـح العيّـنة سـوى من

التحريفات العائدة إلى « عدم الإجابات » وليس من التحريفات المنسوبة إلى التقلّبات العشوائية .

لا يمكن لهاتين الطريقتين الأوليين ، التفريع البعدي واستبدال الأفراد المتخلّفين ، أن تؤدّيا بشكل أكيد إلى تقديرات خالية من التحيّز . كي لا يكون هناك من تحيّز يجب ، في الواقع ، أن يكون في كلّ فرع متوسّط المتغيّرة موضع الدراسة هو نفسه بالنسبة للمجتمع الإحصائي الثانوي المكوّن من الأفراد الذين أجابوا وبالنسبة للمجتمع الإحصائي الثانوي المكوّن من الأفراد الذين لم يجيبوا . بعبارة أخرى ، يجب أن يكون في كلّ فرع استقلالية بين المتغيّرة موضع الدراسة والموقف حيال الحملة . بشكل عام ، لا يمكن تأكيد أيّ شيء بهذا الخصوص .

ج ـ استجواب عيّنة ثانوية تتكوّن من غير المجيبين

وحدها هذه الطريقة هي حقّاً صحيحة وتقود إلى تقديرات خالية من التحيّز .

نعتبر المجتمع الإحصائي مقسوماً إلى مجموعتين ثانويتين :

- المجتمع الإحصائي الثانوي p_1 ، بمقدار N_1 وبمتوسّط m_1 ، مؤلّفاً من الأفراد الذين اخترناهم بالصدفة وأجابوا عن أسئلة الحملة ؛
- المجتمع الإحصائي الثانوي p_2 ، بمقدار N_2 وبمتوسّط m_2 ، مؤلّفاً من الأفراد الذين اخترناهم بالصدفة ولم يجيبوا عن أسئلة الحملة .

طبعاً ، مقدارا هذين المجتمعين الثانويين N_1 وN_2 مجهولان ، وسوف يتمّ تقديرهما بواسطة العددين n_1 وn_2 « للإجابات » و« عدم الإجابات » الملحوظين على العيّنة .

ونعيّن بين غير المجيبين الـ n_2 ، بواسطة سحب مستنفِد ، عيّنة ثانوية تتكوّن من n_2' فرداً نقوم باللازم كي نحصل منهم على إجابة .

نقدّر متوسّط المجتمع الإحصائي :

$$m = \frac{N_1}{N} m_1 + \frac{N_2}{N} m_2$$

بواسطة :

$$\bar{y}_a = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}_2'$$

حيث $\bar{x}_2'$ ترمز إلى متوسّط المتغيّرة X الملحوظ على العيّنة الثانوية . هذا التقدير هو غير متحيّز .

القسم III

كيف نضع خطّة للبحث الإحصائي ؟
مثال : خطّة بحث حملات الـ I.N.S.E.E [1]

1 . الدرجة الأولى من البحث: A . التفريع ؛ B . سحب الوحدات الأوّلية
2 . الدرجة الثانية من البحث . ـ 3 . الدرجة الثالثة من البحث .

على الصعيد العملي ، تتناول خطّة البحث الإحصائي معظم المناهج التي درسناها خلال هذا الفصل وقد تأخذ لهذا السبب شكلاً معقّداً كثيراً .

سوف نعرض تنظيم خطّة للبحث الإحصائي بأخذنا كمثل خطّة حملات الـ I.N.S.E.E [2] .

من أجل معظم الحملات التي يقوم بها حول الأسر ، يعتمد الـ I.N.S.E.E ، في الواقع ، نفس خطّة البحث الإحصائي . والعيّنات ، التي يتغيّر حجمها مع الحملات (من 5000 إلى 20000 مسكن ، بشكل عام) ، هي عيّنات من المساكن ، ويخضع للحملة كلّ الأشخاص الذين يقيمون عادة في المساكن المعيّنة . تتألّف قاعدة البحث الإحصائي من سجلّ شهادات السكن المنبثق عن الإحصاء الأخير (أحدث إحصاء) ، ويُكمّل هذا السجلّ بلائحة مساكن « جديدة » أنجزت منذ ذلك الحين .

تتحقّق الحملات بواسطة « مقابلات » يجريها باحثون مؤهّلون خصّيصاً . وتستبعد ضرورة تخفيض كلفة التنقّل تحقيق بحث إحصائي نموذجي لتفسح المجال أمام بحث على عدّة درجات . في الحقيقة تُصاغ خطّة البحث نفسها بشكل يمكن معه الاحتفاظ بنفس الوحدات الأوّلية خلال عدد معينّ من السنوات يسمح بتجنيد باحثين محلّيين وتخفيف نفقات الإستيفاء اليومي لسجلّ المساكن (مراقبة إنجاز المساكن الجديدة في الوحدات الأوّلية المعيّنة إنطلاقاً من رخص البناء) .

خطة البحث الإحصائي هي على ثلاث درجات :

ـ الدرجة الأولى . ـ سحب عيّنة من الوحدات الأوّلية . هذه الوحدات هي إمّا وحدات

(1) Institut National de la Statistique et des Etudes Economiques
للمعهد الوطني للإحصاء والدراسات الإقتصادية .

(2) المرجع ف. شارتيه F. Chartier ، خطّة البحث الإحصائي لحملات الـ I.N.S.E.E حول الأسر منذ 1969 ، باريس ، I.N.S.E.E .

مدينية (مدن منفردة أو تجمّعات متعدّدة القرى أو النواحي) ، إمّا نواح ريفية متجمّعة في مقاطعات . عيّنة الوحدات الأوّلية هي نفسها لكلّ الحملات وتحفظ خلال عدّة سنوات : إنّها العيّنة الرئيسة .

- الدرجة الثانية .ـ سحب عيّنة من النواحي . النواحي التي نأخذها من الوحدات المدينية ونعيّنها تُستعمل هي أيضاً لعدّة سنوات ، بينما النواحي الريفية التي نعيّنها تُجدّد عند كلّ حملة بحكم أحجامها الصغيرة .

- الدرجة الثالثة .ـ سحب عيّنة من المساكن . تُسحب المساكن المعيّنة خصّيصاً لكلّ حملة ، وتؤخذ احتياطات خاصّة كي لا يمكن لنفس المسكن أن ينتمي إلى عيّنات تتعلّق بحملات مختلفة . وحسب هدفها وموضوعها تجرى الحملة إمّا على الأسر إمّا على الأفراد ، وفي هذه الحالة الأخيرة يكون المسكن عبارة عن عنقود من الأفراد .

في الدرجتين الأولى والثانية من البحث الإحصائي نجري السحوبات بعد أن نفرّع . ولا وجود للتفريع بشكل عام ، عند الدرجة الثالثة .

1 . الدرجة الأولى من البحث الإحصائي

A . التفريع

قبل السحب ، نفرّع الوحدات الأوّلية في آن واحد حسب كبر المنطقة والفئة . المناطق الكبيرة ، وعددها 8 ، هي الـ Z.E.A.T. (مناطق دراسات وتنظيم الأقاليم في البلد) :

المنطقة الباريسية .

الحوض الباريسي : Champagne ، Picardie ، Haute-Normandie ، Basse-Normandie ، Centre ، Bourgogne ؛

الشمال ؛

الشرق : Lorraine ، Alsace ، Franche-Comté ؛

الغرب : بلاد اللوار ، Bretagne ، Poitou-charentes ؛

الجنوب الغربي : Aquitaine ، Midi-Pyrénées ، Limousin ؛

الوسط الشرقي : Rhône-Alpes ، Auvergne ؛

المتوسّط · Languedoc-Roussillon ، Provence-Côte d'Azur ، كورسيكا .

فئات الوحدات الأوّلية ، وعددها 5 ، هي :

نواحي المقاطعات الريفية

- الريفية كلّياً : الفئة 0
- الريفية جزئياً : الفئة 1 ؛

ـ الوحدات المدينية من :

- أقلّ من 20000 نسمة : الفئة 2
- من 20000 إلى أقلّ من 100000 نسمة : الفئة 3
- 100000 نسمة وما فوق : الفئة 4 .

إذن ، يوجد ما مجموعه (40 = 4 × 8) 40 فرعاً . يعطينا الجدول 29 توزيع الوحدات الأوّلية بين الفروع تبعاً لعدد المساكن في إحصاء 1968 (أمكنة الإقامة الرئيسية ، المساكن الشاغرة وأمكنة الإقامة الثانوية) .

الجدول 29 . توزيع المساكن المحصيّة (بالآلاف) والوحدات الأوّلية بين الفروع .

منطقة Z.E.A.T.	فئة الوحدة الأوّلية 4		3		2		1		0		المجموع	
	و.أ	المساكن	و.أ	المساكن	و.أ	المساكن	و.أ	المساكن	و.أ	المساكن	و.أ	المساكن
المنطقة الباريسية	1	3 176	2	117	3	136	2	113	1	40	9	3 582
الحوض الباريسي	9	582	10	551	12	631	15	756	16	810	62	3 330
الشمال	9	690	3	166	4	188	3	140	1	36	20	1 220
الشرق	8	471	4	223	7	330	6	316	4	210	29	1 550
الغرب	6	404	6	347	8	397	10	542	12	632	42	2 322
الجنوب الغربي	5	477	5	264	6	312	6	323	11	591	33	1 967
الوسط الشرقي	4	687	7	400	6	335	8	390	7	377	32	2 189
المتوسّط	8	952	6	327	6	311	5	251	5	270	30	2 111
فرنسا	50	7 439	43	2 395	52	2 640	55	2 831	57	2 966	257	18 271
متوسّط عدد المساكن في كلّ و. أ معيّنة				55,7		50,8		51,5		52,0		

و. أ تعني وحدة أوّلية .

B . سحب الوحدات الأوّلية

الوحدات الـ 50 المدينية التي تتكوّن من 100000 نسمة وأكثر (الفئة 4) تؤلّف وحدات أوّلية كبيرة ومتنوّعة بما فيه الكفاية : إذ تؤخذ كلّها في العيّنة ، وتعطي كلّ

منها عدداً من المساكن المعيّنة يتناسب مع العدد الإجمالي للمساكن التي تنتمي إليها . يمكننا إذن الاعتبار أنّ كلاً من هذه الوحدات المدينية تؤلّف فرعاً خاصّاً تكون فيه درجة البحث الإحصائي الأولى مستنفِدة .

في الفروع الأخرى (الفئات من 0 إلى 3) ، نعين الوحدات الأوّلية بواسطة سحب منهجي ، مع احتمالات تناسبية مع أحجامها (عدد المساكن المحصيّة) ، بواقع وحدة أوّلية معيّنة لكل 50000 مسكن تقريباً . نثبّت أوّلاً عدد الوحدات الأوّلية التي علينا سحبها (العدد المدوّن في الجدول 29) ، ثم نحدّد أساس متوالية السحب الحسابية : عدد مساكن الفرع / عدد الوحدات الأوّلية المعيّنة .

مثلاً . الحوض الباريسي . الفئة 3 من الوحدات الأوّلية (وحدات مدينية من 20000 إلى أقلّ من 10000 نسمة) .

العدد الإجمالي لمساكن الفرع هو 550773 . تقودنا قاعدة الـ 50000 مسكن إلى أخذ 10 وحدات أوّلية للعيّنة ، وأساس متوالية السحب الحسابية هو : 550773/10=55077 . ونأخذ كقاعدة لهذه المتوالية عدداً نسحبه بالصدفة بين 1 و55080 (3+55077 ، حيث 3 هي باقي القسمة السابقة) . وتصبح وحدات العيّنة الأوّلية هي الوحدات التي نعيّنها ، حسب طريقة حواصل الجمع المتراكمة (أنظر سابقاً ، القسم I ، الفقرة 3.B ، ص 315) ، بواسطة عناصر هذه المتوالية الحسابية المختارة بالصدفة .

ويؤدّي سحب الوحدات الأوّلية المنهجي على لوائح منظّمة في كلّ منطقة Z.E.A.T إلى تخصيص كل منطقة بتمثيل تقريباً تناسبي مع حجمها .

2 . الدرجة الثانية من البحث الإحصائي

إنّ الوحدات المدينية التي عيّناها عند الدرجة الأولى هي إمّا مدن منفردة (ناحية واحدة) ، إمّا تجمّعات متعدّدة النواحي .

عند الدرجة الثانية ، نأخذ المدن المنفردة بكاملها في العيّنة (بحث إحصائي مستنفِد) . أمّا التجمّعات متعدّدة النواحي ، المؤلّفة أكثر الأحيان من مدينة ـ مركز ومن نواحٍ أصغر ، فنفرّعها بعد فصلنا المدينة ـ المركز . نأخذ هذه الأخيرة بكاملها في العيّنة ، فيما نسحب بعض النواحي الأخرى بالصدفة مع احتمالات تناسبية مع حجمها . ويجري توزيع مساكن العيّنة بين المدينة ـ المركز والنواحي الأخرى تناسبياً مع العدد الإجمالي لمساكنها .

مثلاً . المنطقة البـاريـسية ، الفئة 3 من الوحدات الأوّلية (الوحـدات المدينيـة من 20000 إلى أقلّ من 100000 نـسمة) .

معدّل البحث الإحصائي : t = 1/2000 .

العدد الإجمالي لمساكن الفرع هـو 116844 . قاعـدة المسـاكن الـ 50000 تؤدّي إلى أخذ وحدتي عيّـنة أوّليتين .

يجب أن يكون عدد مساكن العيّـنة : 58=(1/2000)×116844 لمجمـوع الفرع ، و29=58/2 بالنسبة لوحدة عيّنة أوّلية .

لنفتـرض أنّـه عند الـدرجة الأولى ، كـان تجمّـع Mantes-la-jolie واحـدة من الـوحـدتـين الأوّليتـين المعيّـنتـين . يتضمّـن هـذا التجمّـع مـدينـة ـ مـركـزاً (Mantes-la-Jolie) و8 نواح أخرى (الجدول 30) .

الجدول 30 . مثل عن اختيار نواحي العيّـنة :

تجمّـع Mantes-la-jolie

الفروع الثانوية	مقدار الفرع الثانوي	عدد المساكن المعيّـنة في الفرع الثانوي	النواحي	مقدار الناحية
1	8 979	14	Mantes-la-Jolie	8 979
2	9 957	15	Follainville-Dennemont	378
			Gargenville	1 301
			Isson	332
			Limay	2 260
			Porcheville	547
			Buchelay	328
			Magnanville	107
			Mantes-la-Ville	4 704
المجموع	18 936	29		18 936

لنفترض أنّنا ، لأسباب تتعلّق بسعر التكلفة ، وضعنا قاعدة تفرض على أن لا يقلّ عدد مساكن العيّنة في كلّ ناحية عن 10 .

ضمن هذه الشروط ، سوف نقطّع التجمّـع إلى فرعين ثانويين اثنين :

1 . المدينة ـ المركز ؛

2 . النواحي الأخرى .

الناحية Mantes-la-Jolie التي تكوّن الفرع الثانوي 1 ، ندخلها بكاملها في العيّنة مع (14=(8979/18936)×29) 14 مسكناً معيّناً .

ونسحب من ضمن نواحي الفرع الثانوي 2 ، واحدة بالصدفة نأخذ منها (15=(9957/18936)×29) 15 مسكناً معيّناً .

في الوحدات الأوّلية المعيّنة المؤلّفة من النواحي الريفية المجمّعة مقاطعات ، نعمد إلى سحب منهجي باحتمالات تناسبية مع أحجامها لـ 2 ، 3 أو 4 نواح ـ عيّنة . ويتغيّر عدد نواحي العيّنة حسب عدد المساكن المعيّنة (الذي يتعلّق بدوره بمعدّل البحث الإحصائي) ومتوسّط حجم نواحي المقاطعة (كون حجم النواحي يختلف بشكل ملحوظ من منطقة إلى أخرى) .

3 . الدرجة الثالثة من البحث الإحصائي

في معظم الحملات الإحصائية ، وبحكم درجات البحث المتالية ، يكون معدّل البحث الإحصائي النهائي نفسه مهما كان الفرع . بما أنّ سحب وحدات البحث عند الدرجتين الأولى والثانية قد تمّ باحتمالات تتناسب مع أحجامها ، فإنّ عدد المساكن المعيّنة (المحسوب بشكل يراعي معدّل البحث النهائي) هو نفسه في كلّ ناحية معيّنة من نفس الوحدة الأوّلية . العيّنة هي إذن مرجّحة بذاتها (أنظر القسم I ، الفقرة 4.C ، ص 322) .

يجري تعيين مساكن العيّنة بواسطة سحب منهجي على لوائح بالمساكن موضوعة للنواحي المعيّنة ، وغالباً بدون تفريع محتمل . ولكن بما أنّ لوائح المساكن هي مصنّفة حسب الشوارع والأحياء داخل النواحي ، فإنّ السحب المنهجي يضمن للعيّنة توزيعاً جغرافياً مرضياً .

إلّا أنّه قد يحدث في بعض الحملات أن يكون معدّل البحث الإحصائي مختلفاً ، مثلاً حسب فئة الوحدة الأوّلية أو فئة ربّ الأسرة الاجتماعية ـ المهنية . في الحالة الأولى ، يحدّد عدد المساكن المعيّنة في كلّ فرع ، وفي الثانية ، بافتراضنا أنّنا نريد مثلاً سحب عيّنة بمعدّل 1/2000 للفئات الاجتماعية ـ المهنية 2 ، 3 أو 4 (أرباب العمل والكوادر) وبمعدّل 1/4000 للفئات الأخرى ، نسحب عيّنة متجانسة بالمعدّل الأعلى (أي 1/2000) . بعد ذلك نحذف واحداً على اثنين من المساكن المعيّنة التي تشغلها أسر تكون فئة ربّها الاجتماعية ـ المهنية مختلفة عن 2 ، 3 أو 4 . إنّ هذا المنهج يحفظ لسحوبات لاحقة الصفة التمثيلية للمساكن التي تبقى على اللائحة بعد اختيار العيّنة المتجانسة بمعدّل 1/2000 .

الفصل الثامن

تحليل السلاسل الزمنية

السلسلة الزمنية هي سلسلة من المشاهدات المرتّبة تبعاً للوقت : مقدار السكّان السنوي ، القيمة السنوية للانتاج الوطني الخام ، المستوى الشهري لمؤشّر الأسعار ، مجموع المبيعات الشهري لشركة معيّنة ، عدد أجزاء مؤسّسة معيّنة آخر كلّ شهر ، الخ .

لقد كان الوصف العام للسلاسل الزمنية وبشكل خاص تمثيلها البياني موضوعاً عرض في كتاب « الإحصاء الوصفي » ، الفصل III ، القسم III .

إنّ دورية المشاهدات متغيّرة : أكثر الأحيان تكون السلاسل الزمنية شهرية ، فصلية أو سنوية . وأحياناً هي أسبوعية ، يوميّة وحتّى بالساعة (دراسة حركة المرور ، الخطّ الهاتفي) أو ، بالعكس ، كلّ ستين أو كلّ عشر سنوات (مثلاً ، إحصاءات السكّان في العديد من البلدان) .

أن نعطي حكماً على تطوّر حديث لسلسلة زمنية معيّنة ليس ، بشكل عام ، مهمّة سهلة . فعدد لا بأس به من السلاسل يقدّم في الواقع تغيّرات دورية على درجات متفاوتة من الانتظام تفسّر تأثير عوامل مثل الإجازات السنوية ، الفصل أو العادات . يمكن لهذه التغيّرات الموسمية أن تقنّع تطوّر الظاهرة الحقيقي ، وكي نبرز هذا التطوّر ، من الضروري أن نحلّل السلسلة وأن نفصل العامل الموسمي عن بقيّة المكوّنات : مثلاً ، يجري تشخيص ميول تطوّر مؤشّر الإنتاج الصناعي بعد تصحيح

التغيّرات الموسمية . إن I.N.S.E.E [1] ينشر بانتظام عدداً كبيراً من السلاسل بتغيّرات موسمية مصحّحة .

القسم I

صورة التحليل

1 . مكوّنات سلسلة زمنية : A . تغيّرات الاتجاه العام أو trend ؛ B . الحركة الدورية ؛ C . التغيّرات الموسمية ؛ D . التغيّرات العرضية أو المتبقّية ؛ E . فائدة تصحيح التغيّرات الموسمية .ـ 2 . نماذج التكوين : A . الصورة الجمعية ؛ B . الصورة المضاعفة .ـ 3 . طرق التجزئة : A . التحليلية ؛ B . الطرق التجريبية .

يمكننا تجزئة السلاسل الزمنية إلى عدّة عناصر قابلة لأن تتّحد حسب نماذج مختلفة .

1 . مكوّنات سلسلة زمنية

بشكل عام ، يمكننا أن نميز في تطوّر سلسلة زمنية ، أربع مكوّنات .

A . تغيّرات الاتجاه العام أو Trend

يمثّل الـ trend تطوّر الظاهرة العام لمدى طويل ، مرتبطاً بالنمو العام للاقتصاد : انخفاض عدد العاملين في الزراعة ، تزايد الانتاج الصناعي ، تزايد استهلاك الكهرباء ، على سبيل المثال .

B . الحركة الدورية

حول الاتجاه العام يوجد تقلّبات تتعلّق بالتغيّرات الظرفية وبصورة خاصّة بتتابع مراحل الدورة الإقتصادية : ازدهار ، أزمة ، انحطاط ، نهضة .

في فرنسا مثلاً خلال الأعوام الأخيرة تمكّن المعدّل السنوي لتزايد الانتاج الصناعي أن يبلغ 12% في فترة الازدهار والتغى (عادل صفراً) في فترة الانحطاط ، بينما متوسّط المعدّل الذي يمثّل الاتجاه العام هو تقريباً 6,5% في السنة .

---

(1) المعهد الوطني للإحصاء والدراسات الاقتصادية .

لقد شكّلت التغيّرات الدورية موضع الكثير من نظريات علماء الاقتصاد ، وبما أنّ هذه المسألة المناقشة كثيراً تخرج عن إطار اهتمامات هذا الكتاب العملية ، لن نحاول الفصل بين trend وحركة دورية ، وسوف نرمز إليهما سويّاً باسم « الحركة غير الموسمية » أو أحياناً الحركة الظرفية .

C . التغيّرات الموسمية

التغيّرات الموسمية هي التقلّبات الدورية المنتظمة قليلاً أو كثيراً والتي تتصادف مع الحركة غير الموسمية . وقد تكون دورتها يومية (حركة المرور كلّ ساعة) ، أسبوعية (عدد ساعات العمل اليومية) أو سنوية (المؤشّر الشهري للانتاج الصناعي ، مجموع المبيعات الشهري في المخازن الكبيرة) . وهي متعدّدة الأسباب ، دورة الفصول ، طريقة الحياة ، العادات ، الأحكام القانونية ، الخ . ، تحدث تأثيراتها بشكل ملحوظ عند تاريخ محدّد . من أهمّها نذكر :

- الإجازات : تُترجَم الإجازات السنوية كلّ صيف بيطء ملحوظ في النشاط وبهبوط في معظم الكمّيات الاقتصادية الرئيسية . بصورة خاصّة ، يسجّل الانتاج الصناعي فراغاً موسمياً كبيراً .

- التفاوت في عدد أيام الأشهر المختلفة : تتأثّر معظم النشاطات الاقتصادية بعدد أيام العمل في كلّ شهر . فتنقّل الأعياد غير الثابتة وتوقّف العمل لسبب معيّن بشكل ملائم أو غير ملائم يمكنه أن يجعل مقارنة الشهر نفسه بين سنتين متاليتين عسيرة . ونلجأ في بعض السلاسل ، بصورة خاصّة مؤشّرات الانتاج الصناعي ، إلى إدخال تصحيح في عدد أيام العمل ، يسبق التصحيح المسمّى بتصحيح التغيّرات الموسمية البحت .

- العوامل المناخية : عند تحليل أخير ، تظهر هذه العوامل كمصدر لمعظم التغيّرات الموسمية ؛ الإجازات السنوية ، مثلاً ، تؤخذ في الصيف بشكل عام لأنّه الفصل الأنسب . ولكن في بعض الحالات يكون تأثير العوامل المناخية مباشراً أكثر : البرد القارس يبطىء نشاط البناء إلى حدٍ بعيد؛ الحرارة تؤثّر في استهلاك الكهرباء من قبل الأفراد (تدفئة ، تكييف) . أكثر الأحيان تؤثّر العوامل المناخية في الظواهر الإقتصادية بطرق معقّدة : تؤثّر بشكل خاص عن طريق عرض وطلب بعض السلع .

- دورية عرض وطلب بعض المنتوجات : غالباً ما تأتي هذه الدورية ، من ناحية

العرض ، نتيجة الإيقاع الموسمي للانتاج الزراعي . أمّا تنظيم الأسواق وإمكانيات التخزين فلا تضبط إلاّ بشكل ناقص تغيّر وفرة وأسعار هذه المنتوجات خلال العام . من ناحية الطلب ، يُسجّل أيضاً بالنسبة لبعض السلع تغيّرات منتظمة بعض الشيء : مبيعات آخر السنة ، طلب السيّارات في الربيع ، الخ .

D . التغيّرات العرضية أو المتبقّية

يحدث حول الحركة المذكورة سابقاً بعض التقلّبات العشوائية . وهي تعود إمّا إلى عدد كبير من الأسباب الصغيرة ـ عندها يكون مدى التقلّبات ضعيفاً بشكل عام ـ إمّا إلى تدخّل حوادث طارئة : إضراب ، انهيار مالي ، تعديل في القانون الضريبي ، الاجتماعي أو الاقتصادي ، الخ . هذه التغيّرات تمثّل في تطوّر السلسلة ناحية لا يمكن للمكوّنات السابقة أن تأخذها بعين الاعتبار . لهذا السبب نعطيها أحياناً إسم التقلّبات المتبقّية .

E . فائدة تصحيح التغيّرات الموسمية

سوف نثبت ، على مثل بالأرقام ، ضرورة تصحيح التغيّرات الموسمية لتفسير تطوّر سلسلة معيّنة .

لتسهيل الأمر ، سوف نتصوّر سلسلة خالية من التقلّبات العرضية ونفترض أنّ المعطيات الملحوظة في عامي 1969 و1970 هي حاصل جمع الحركة غير الموسمية والمظهر الموسمي العام التاليين :

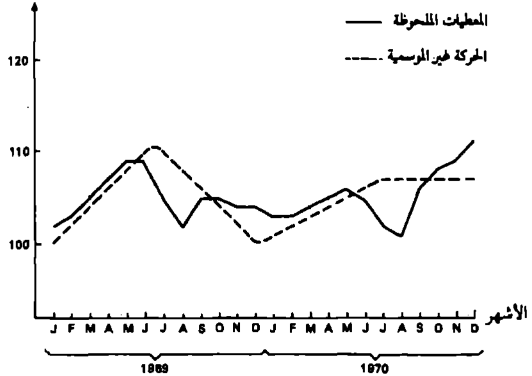
	الحركة غير الموسمية		المظهر الموسمي العام	المعطيات الملحوظة		تغيّرات 1970 بالنسبة للشهر المطابق من العام 1969
	1969 (1)	1970 (2)	(3)	1969 (4)	1970 (5)	(6)
كانون الثاني	100	101	+ 2	102	103	+ 1 %
شباط	102	102	+ 1	103	103	0
آذار	104	103	+ 1	105	104	− 1 %
نيسان	106	104	+ 1	107	105	− 2 %
أيار	108	105	+ 1	109	106	− 3 %
حزيران	110	106	− 1	109	105	− 4 %
تموز	110	107	− 5	105	102	− 3 %
آب	108	107	− 6	102	101	− 1 %
أيلول	106	107	− 1	105	106	+ 1 %
تشرين الأوّل	104	107	+ 1	105	108	+ 3 %
تشرين الثاني	102	107	+ 2	104	109	+ 5 %
كانون الأوّل	100	107	+ 4	104	111	+ 7 %

خلال هذين العامين ، تبقى الحركة الموسمية كما هي تماماً (الشكل 61) .

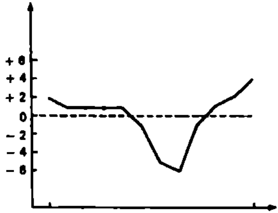
التغيّر الحقيقي للسلسلة تعرضه الحركة غير الموسمية : تزايد الكمّية من كانون الثاني (يناير) إلى حزيران (يونيه) 1969 ، هبوط من تمّوز (يوليو) إلى كانون الأوّل (ديسمبر) ، ثمّ تزايد من جديد حتّى تمّوز (يوليو) 1970 يتبعه استقرار (الشكل 60) .

لننسَ الآن كلّ ما نعرفه ولنفترض ، كما سيكون الحال فعلاً ، أنّنا لا نملك سوى المعطيات الملحوظة . عند رؤية هذه المعطيات ، من الصعب جدّاً أن نتبيّن اتجاهات تطوّر السلسلة . بالنسبة للعام 1970 مثلاً ، قد نعتقد بوجود هبوط انطلاقاً من حزيران ثمّ نهضة سريعة في آخر السنة ، بينما الاتجاهات الحقيقية هي تزايد حتّى شهر تمّوز يتبعه استقرار .

أمام هذه الصعوبات ، هناك حلّ معتمد غالباً يقوم على تقريب معطيات شهر معيّن مع معطيات الشهر المطابق من السنة السابقة ، وذلك شهراً فشهراً . بهذا العمل ، نفكّر بحذف فعل التغيرات الموسمية . في الواقع ، وكما نرى على مثلنا ، قد يقودنا



الشكل 60 . تمثيل السلسلة بيانياً .



الشكل 61 . المظهر الموسمي العام

هذا المنطق إلى نتائج خاطئة .

يعرض العامود (6) من الجدول السابق مقارنة منهجية للمعطيات التابعة لنفس الشهر . وتوحي النتائج الحاصلة بالتفسير التالي : من كانون الثاني إلى حزيران 1970 تمضي نسب التغيّر المئوية متناقصة ، إذن يتّجه الوضع إلى التدهور ؛ بالمقابل ، انطلاقاً من شهر تمّوز تبدأ نسب التغيّر المئوية بالتزايد مسرعة ، بالتالي سرعان ما يتحسّن الظرف . وقد يزيد من قوّة هذا الحكم المفعول النفسي للأعداد الإيجابية : انطلاقاً من أيلول يبدو الوضع جيّداً لأنّنا نلاحظ في هذه الأشهر مستوى أعلى من مستوى السنة السابقة . هذا التشخيص هو خاطىء كلّياً ؛ في الحقيقة الوضع جيّد حتّى شهر تمّوز (تزايد منتظم) ، ثمّ نسبياً غير ملائم (ركود) . « هذا لأنّنا ننسى خلال اتّباعنا لهذا النمط من التفكير أنّ تطوّر النسب المئوية موضع السؤال يتعلّق ، ليس فقط بمظهر السنة الجارية ، بل أيضاً بمظهر السنة السابقة »[1] . : مقارنة مع الهبوط في الفصل الثاني من عام 1969 ، يبدو الركود المطابق في العام 1970 تحسناً .

إذن ، يقوم الحلّ الصحيح على مقارنة تطوّر السنة الجارية ، ليس فقط مع السنة السابقة ، بل مع مجموع السنوات السابقة . بعد تحديد النموذج المناسب لتكوين الحركة الظرفية والحركة الموسمية ، نحدّد « متوسّط الجانبية الموسمية » انطلاقاً من سلسلة مشاهدات تتعلّق بعدد كبير من السنوات . عندئذٍ يكفي مقارنة تطوّر سنة معيّنة مع هذه الجانبية النظرية كي نخلص إلى الاتجاهات الحقيقية للسلسلة .

---

(1) J. Méraud ، A. Tymen و R. Jaulent ، التغيّرات الموسمية للنشاط الاقتصادي : طريقة تحليل ، تطبيق على الانتاج الصناعي وتشغيل اليد العاملة ، دراسات ووقائع ، نيسان 1960 .

2 . نماذج التكوين

حيث لا نحاول فصل الاتّجاه العام عن الدورة ، نجد ثلاث مكوّنات :

- المكوّنة غير الموسمية أو الظرفية ،
- المكوّنة الموسمية ،
- التغيّرات العشوائية .

إنّ تجزئة السلسلة إلى هذه المكوّنات الثلاث تفترض عدداً من الفرضيات تتعلّق بكيفية تركيب وطبيعة هذه المكوّنات .

لنرمز بالنسبة للشهر j من السنة i بواسطة :

y_{ij} إلى القيمة الملحوظة للسلسلة الزمنية ؛

c_{ij} إلى قيمة المكوّنة الظرفية ؛

s_{ij} إلى المكوّنة الموسمية ؛

e_{ij} التغيّرات المتبقية أو العرضية .

تمثّل السلسلة الزمنية عامّة على جدول مزدوج المدخل. أنظر الجدول 31 :

على السطر ، نجد السنة (الدليل السفلي i) ،

على العامود ، نجد الشهر (الدليل السفلي j) .

وقد اخترنا هذا الوضع كي نسهّل عرضنا في الحالة التي نصادفها تكراراً وهي حالة سلسلة شهرية بتغيّرات موسمية على حقبة سنوية . بشكل عام ، تتطابق حقبة سلسلة زمنية مع دورة كاملة للتغيّرات الموسمية . وهذه الدورة قد تكون مثلاً اليوم في حالة تحليل التغيّرات كلّ ساعة لخطوط الهاتف ؛ وهي السنة أكثر الأحيان بالنسبة لدراسة الظواهر الاقتصادية .

		« الأشهر »				
	$i \backslash j$	1	2 ...	j ...		p
« السنوات »	1	y_{11}	y_{12} ...	y_{1j} ...		y_{1p}
	2	y_{21}	y_{22} ...	y_{2j} ...		y_{2p}
	⋮	⋮				
	i	y_{i1}	y_{i2} ...	y_{ij} ...		y_{ip}
	⋮	⋮				
	n	y_{n1}	y_{n2} ...	y_{nj} ...		y_{np}

الجدول 31 .
تمثيل سلسلة زمنية

الدليل السفلي i يعاين رقم الحقبة أو الدورة :

$$i = 1, 2, \ldots, n$$

والدليل السفلي j يعاين ، داخل الدورة i ، تاريخ نقل الملاحظة :

$$j = 1, 2, \ldots, p$$

مثلاً داخل الدورة السنوية ، التواريخ قد تكون الأشهر (p=12) ، أو الفصول (p=4) أو الأسابيع (p=25) .

يمكننا أيضاً أن نعاين الملاحظة مباشرة بواسطة دليل تاريخ الملاحظة أو المشاهدة t داخل السلسلة . إذا كانت هذه السلسلة تتعلّق بعدد صحيح من الدورات ، يكون لدينا :

$$t = p(i-1) + j$$

بصورة خاصّة ، في حالة السلسلة الشهرية :

$$t = 12\,(i-1) + j$$

إذن نرمز إلى مشاهدة معيّنة بلا تمييز بواسطة :

y_t أو y_{ij}

أبسط نماذج تركيب العناصر المكوِّنة لسلسلة زمنية هما الصورتان الجمعية أو المضاعِفة .

A . الصورة الجمعية

تفترض الصورة الجمعية :

$$y_t = c_t + s_t + \varepsilon_t$$

أنّ المكوِّنة الموسمية للسلسلة ، كما التغيّر المتبقّي ، هما مستقلّان عن الحركة غير الموسمية .

B . الصورة المضاعِفة

تفترض الصورة المضاعِفة : $y_t = c_t + c_t s_t + \varepsilon_t = c_t(1 + s_t) + \varepsilon_t$

أنّ المكوِّنة الموسمية ، الممثّلة بواسطة $c_t s_t$ ، هي تناسبية مع الحركة الظرفية .

ونستعمل في بعض الحالات شكلاً آخر للصورة المضاعِفة ، نفترض فيها أنّ

التغيّر المتبقّي يتناسب بدوره مع حاصل جمع المكوّنتين الأوليين :

$$y_t = c_t(1 + s_t) + c_t(1 + s_t)\,\varepsilon_t = c_t(1 + s_t)(1 + \varepsilon_t)\,. \qquad (3)$$

لنأخذ لوغاريتم عنصري هذه العبارة :

$$\log y_t = \log c_t + \log(1 + s_t) + \log(1 + \varepsilon_t)\,.$$

إذا وضعنا :

$$Y_t = \log y_t\,, \qquad C_t = \log c_t\,, \qquad S_t = \log(1 + s_t)$$

وإذا لاحظنا أنّ :

$$\log(1 + \varepsilon_t) \approx \varepsilon_t\,,$$

كوننا اعتبرنا ε_t صغيراً ، فإنّ هذا الشكل الثاني يتحوّل إلى الصورة الجمعية :

$$Y_t = C_t + S_t + \varepsilon_t\,.$$

3 . طرق التجزئة

إذن ، تقوم تجزئة السلسلة الزمنية على تقدير قيمتي المكوّنة الظرفية c_t والمكوّنة الموسمية s_t ، وذلك لكلّ تاريخ مشاهدة .

تُستعمل لهذه الغاية فئتان رئيسيتان من الطرق : الطرق التحليلية والطرق التجريبية .

A . الطرق التحليلية

في هذا النوع من الطرق ، نضع فرضية حول الشكل التحليلي للمكوّنتين الظرفية والموسمية .

نضع مثلاً الفرضيتين التاليتين :

- الحركة الظرفية هي دالّة خطّية تبعاً للوقت : $C_t = at + b\,.$

- الحركة الموسمية هي دالّة دورية تماماً ، دورتها p=12 ، وتأخذ القيم γ_j $(j = 1, 2, \dots, 12)$:

$$s_t = s_{ij} = \gamma_j$$
$$s_{t+p} = s_{i+1,j} = \gamma_j$$
الخ

ضمن فرضية صورة جمعية للتكوين :

$$y_t = c_t + s_t + e_t$$

وإذا استبدلنا c_t وs_t بشكلهما التحليلي نحصل على :

$$y_t = at + b + \gamma_j + e_t \, .$$

وإذا وضعنا :

$$b + \gamma_j = b_j$$
$$y_t = at + b_j + e_t \, .$$

بهذا العمل نكون قد حدّدنا « نموذجاً » لتطوّر السلسلة الزمنية . المسألة تصبح إذن مسألة تقدير المتغيّرين الوسيطيين a وb_j $(j=1, 2, .., u)$ في هذا النموذج بشكل تكون معه « المسافة » بين القيم الملحوظة y_t والقيم النظرية $at+b_j$ أضعف ما يمكن . بشكل عام ، نحدّد هذه المسافة بحاصل جمع مربّعات البواقي e_t ونبحث عن القيم a وb_j التي تجعلها حدّاً أدنى (طريقة المربّعات الصغرى) .

إذن يبدو تحليل السلاسل الزمنية بواسطة الطريقة التحليلية كحالة خاصّة من مسألة تسوية دالّة معيّنة مع سلسلة مشاهدات (راجع التسوية الخطّية في الفصل IV ، القسم III ، خاصّة الفقرة 1.C) .

تقدّم الطريقة التحليلية حسنات عديدة ، إنّها تتمتّع بشكل خاص بأسس نظرية متينة وتسمح بتقييم تباين المتغيّرات الوسيطية المقدّرة ، أي بحساب دقّة تقدير مختلف مكوّنات السلسلة . ولكنّها تشكو من عيب كبير ، وهو أنّه لا يمكن تطبيقها إلاّ على سلاسل نستطيع تمثيلها بشكل صحيح بواسطة دالّة تحليلية : دالّة خطّية ، دالّة أسّية ، ذو الحدود ، الخ . ولكنّنا نعرف أنّه بالنسبة لمعظم السلاسل الزمنية المتعلّقة بالظواهر الاقتصادية ، لا يسمح لنا مسلك المكوّنة غير الموسمية بأخذ صور تطوّر بهذه السهولة .

B . الطرق التجريبية

الطرق التجريبية لا تضع أي فرضية حول مسلك الحركة غير الموسمية . لسوء الحظ لا يمكن عند غياب مرجع إلى نموذج محدّد ، أن نبني طريقة تحليل متينة . وهكذا نلجأ إلى طرق حساب تجريبية . مع هذا ، تُستعمل هذه الطرق بكثرة من أجل تحليل السلاسل الاقتصادية التي نادراً ما تفي بشروط تطبيق الطرق التحليلية : إنّ تحديد شكل

الحركة غير الموسمية والبحث عن المعاملات الموسمية يتمّان بشكل رائج على طريقة المتوسّطات المتحرّكة . ولقد اخترنا أن نعرض هذه الطريقة سهلة التنفيذ وذات التطبيق العام .

القسم II

طريقة المتوسّط المتحرّك

1 . تعريف « المتوسّط المتحرّك » . 2 . خصائص المتوسّط المتحرّك : A . تصفية مكوّنة موسمية دورية ؛ B . تصفية المكوّنة غير الموسمية ؛ C . تصفية التقلّبات العشوائية . 3 . تصحيح التغيّرات الموسمية : A . الفرضيات ؛ B . حساب المعاملات الموسمية ؛ C . مثال تطبيقي : المؤشّر الفصلي للانتاج الصناعي .

1 . تعريف « المتوسّط المتحرّك »

لنأخذ السلسلة الزمنية Y :

$$y_1, y_2, \ldots, y_t, \ldots, y_T .$$

نطلق اسم المتوسّط المتحرّك بطول p للسلسلة Y على العملية التي تحوّل هذه السلسلة إلى سلسلة جديدة Z بواسطة حساب سلسلة المتوسّطات التالية :

$$z_t = z_{l+(p+1)/2} = \frac{1}{p}(y_{l+1} + y_{l+2} + \cdots + y_{l+p})$$
$$= \frac{1}{p}\sum_{i=1}^{p} y_{l+i}, \quad (l = 0, 1, \ldots, T-p)$$

نعطي كلّ متوسّط متال إلى التاريخ t الذي يطابق منتصف الفترة الممتدّة من الحين l+1 إلى الحين l+p :

$$t = \frac{1}{2}[(l+1) + (l+p)] = l + \frac{p+1}{2}.$$

هذا التاريخ الذي يقع في منتصف الدورة يطابق واحداً من تواريخ المشاهدة إذا كان p مفرداً ، ويطابق مركز الفسحة التي تفصل بين تاريخي مشاهدة متاليين ، إذا كان p مزدوجاً .

ونرمز إلى عملية المتوسّط المتحرّك بطول p ، التي نجريها على السلسلة Y ، بواسطة :

$$z_t = M_p(y_t) .$$

مثل 1 . متوسّط متحرّك بطول p=3 :

$$z_t = M_3(y_t)$$

$$z_2 = \tfrac{1}{3}(y_1 + y_2 + y_3)$$

$$z_3 = \tfrac{1}{3}(y_2 + y_3 + y_4)$$

$$\vdots$$

$$z_{T-1} = \tfrac{1}{3}(y_{T-2} + y_{T-1} + y_T) .$$

تطبيق بالأرقام :

t	1	2	3	4	5	6
y_t	103	98	109	111	105	118
z_t	—	103,3	106,0	108,3	111,33	—

بالفعل :

$$z_2 = \frac{103 + 98 + 109}{3} = 103,3$$

$$z_3 = \frac{98 + 109 + 111}{3} = 106,0 \quad \text{etc.}$$

مثل 2 . متوسّط متحرّك بطول p=4 :

$$z_t = M_4(y_t)$$

$$z_{2,5} = \tfrac{1}{4}(y_1 + y_2 + y_3 + y_4)$$

$$z_{3,5} = \tfrac{1}{4}(y_2 + y_3 + y_4 + y_5)$$

$$\vdots$$

$$z_{T-1,5} = \tfrac{1}{4}(y_{T-3} + y_{T-2} + y_{T-1} + y_T) .$$

تطبيق الأرقام :

t	1	2	3	4	5	6	7	8
y_t	101	107	99	109	113	120	107	123
z_t		—	104,00	107,00	110,25	112,25	115,75	—

بالفعل :

$$z_{2,5} = \frac{101 + 107 + 99 + 109}{4} = 104{,}00$$

$$z_{3,5} = \frac{107 + 99 + 109 + 113}{4} = 107{,}00$$

الخ .

نلاحظ أنّ عدداً من القيم ، تطابق طرفي فسحة تغيّر السلسلة الأصلية ، بضيع سدى : لا يمكن تحديد z_1 و z_7 بالنسبة لمتوسّط متحرّك بطول 3 ؛ كذلك $z_{1,5}$ و $z_{7,5}$ بالنسبة لمتوسّط متحرّك بطول 4 ، الخ .

حالة المتوسّطات المتحركة « المزدوجة » $p=2n$

على الصعيد العملي ، غالباً ما نضطر لحساب متوسّطات متحرّكة تتعلّق بعدد مزدوج من الدورات (12 شهراً أو 4 فصول ، مثلاً) . من المزعج أن نحصل بهذه العملية على سلسلة لا تتناسب تماماً مع نفس تواريخ المشاهدة . هكذا رأينا أنّه من المناسب أن نخصّص لتاريخ مشاهدة محدّد المتوسّط الحسابي للمتوسّطين المتحرّكين اللذين يحيطان به . بالتالي نحدّد عملياً متوسّطاً متحرّكاً بطول مزدوج $(p=2n)$ بواسطة :

$$z_t = z_{t+n} = \frac{1}{2}\left[z_{t+n-1/2} + z_{t+n+1/2}\right]$$

$$= \frac{1}{2}\left[\frac{1}{2n}(y_t + y_{t+1} + \cdots + y_{t+2n-1}) + \frac{1}{2n}(y_{t+1} + y_{t+2} + \cdots + y_{t+2n})\right]$$

$$= \frac{1}{2n}\left[\frac{1}{2}y_t + (y_{t+1} + y_{t+2} + \cdots + y_{t+2n-1}) + \frac{1}{2}y_{t+2n}\right].$$

أخيراً ، حسب هذا الاصطلاح ، فإنّ تحديد متوسّط متحرّك بطول مزدوج $(p = 2n)$ يناسب تاريخ المشاهدة t ، يعني حساب المتوسّط المرجّح للـ $2n+1$ مشاهدة التي تحيط بالتاريخ t ، وذلك بتخصيص الوزن :

1/2 إلى المشاهدتين الطرفيتين ،

1 إلى الـ $n-1$ مشاهدة وسيطة .

مثل 3 . متوسّط متحرّك بطول p=4 :

$$
\begin{aligned}
z_3 &= \tfrac{1}{4}[\tfrac{1}{2} y_1 + (y_2 + y_3 + y_4) + \tfrac{1}{2} y_5] \\
z_4 &= \tfrac{1}{4}[\tfrac{1}{2} y_2 + (y_3 + y_4 + y_5) + \tfrac{1}{2} y_6] \\
&\vdots \\
z_{T-2} &= \tfrac{1}{4}[\tfrac{1}{2} y_{T-4} + (y_{T-3} + y_{T-2} + y_{T-1}) + \tfrac{1}{2} y_T] .
\end{aligned}
$$

تطبيق بالأرقام : لنعد إلى المثل السابق

t	1	2	3	4	5	6	7	8
y_t	101	107	99	109	113	120	107	123
z_t	—	—	105,50	108,625	111,25	114,00	—	—

بالفعل :

$$z_3 = \frac{z_{2,5} + z_{3,5}}{2} = \frac{101 + 2(107 + 99 + 109) + 113}{8} = 105{,}50$$

$$z_4 = \frac{z_{3,5} + z_{4,5}}{2} = \frac{107 + 2(99 + 109 + 113) + 120}{8} = 108{,}625$$

الخ .

إنّه المتوسّط المتحرّك المحسوب بمساعدة هذا الاصطلاح هو الذي سنأخذه من الآن فصاعداً بعين الاعتبار ، ونرمز إليه بواسطة :

$$z_t = M_p(y_t) .$$

2 . خصائص المتوسّط المتحرّك

لتبسيط الأمور ، لن نأخذ في هذه الفقرة بعين الاعتبار سوى متوسّطات متحرّكة « مفردة » . بعد ذلك يمكن بسهولة بسط النتائج إلى المتوسّطات المتحرّكة « المزدوجة » ، الموضوعة حسب اصطلاح الحساب المعروض أعلاه . إنّ عمليّة « المتوسّط المتحرّك » هي عبارة عن « مصفاة » : فهي توقف المكوّنات التي تمثّل شكلاً معيّناً وتدع الأخرى تمرّ .

A . تصفية مكوّنة موسمية دورية

لنفترض

$$y_t = c_t + \gamma_t$$

سلسلة زمنية مؤلّفة من مكوّنة ظرفية c_t ومن مكوّنة موسمية $\gamma_t = s_t$، حيث γ_t هي دالّة دورية ودورتها p=2n+1 :

لنحسب المتوسّط المتحرّك بطول p=2n+1 المتعلّق بهذه السلسلة . إنطلاقاً من قاعدة التعريف :

$$z_{t+(p+1)/2} = \frac{1}{p} \sum_{i=1}^{p} y_{t+i}$$

أي ، إذا استبدلنا p بـ 2n+1 :

$$z_{t+n+1} = \frac{1}{2n+1} \sum_{i=1}^{2n+1} y_{t+i} .$$

سوف نضع لتسهيل الرموز :

$$t = l + n + 1 , \qquad k = i - n - 1$$

فنحصل على :

$$z_t = \frac{1}{2n+1} \sum_{k=-n}^{+n} y_{t+k} , \quad (t = n+1, n+2, \ldots, T-n) .$$

إذا استبدلنا y_{t+k} بعبارتها تبعاً لمكوّنتيها :

$$z_t = \frac{1}{2n+1} \sum_{k=-n}^{+n} (c_{t+k} + \gamma_{t+k}) = \frac{1}{2n+1} \sum_{k=-n}^{+n} c_{t+k} + \frac{1}{2n+1} \sum_{k=-n}^{+n} \gamma_{t+k}$$
$$= M_{2n+1}(c_t) + M_{2n+1}(\gamma_t) .$$

المتوسّط المتحرّك لحاصل جمع المكوّنات يساوي حاصل جمع متوسّطاتها المتحرّكة[1] .

لنرمز بواسطة إلى متوسّط المكوّنة الموسمية المتحرّك :

$$\zeta_t = M_{2n+1}(\gamma_t)$$

---

(1) بشكل عام ، المتوسّط المتحرّك هو ، ككلّ متوسّط حسابي ، مؤثّر خطّي :

$$M(x_t + y_t) = M(x_t) + M(y_t)$$
$$M(\lambda x_t) = \lambda M(x_t) .$$

$$\xi_t = \frac{1}{2n+1} \sum_{k=-n}^{+n} \gamma_{t+k}$$

$$\xi_{t+1} = \frac{1}{2n+1} \sum_{k=-n}^{+n} \gamma_{t+k+1}.$$

الفارق :

$$\xi_{t+1} - \xi_t = \frac{1}{2n+1} \sum_{k=-n}^{+n} (\gamma_{t+k+1} - \gamma_{t+k})$$

وإذا وسّعنا :

$$\xi_{t+1} - \xi_t = \frac{1}{2n+1} [(\gamma_{t-n+1} - \gamma_{t-n}) + (\gamma_{t-n+2} - \gamma_{t-n+1}) + \cdots + (\gamma_{t+n} - \gamma_{t+n-1})$$
$$+ (\gamma_{t+n+1} - \gamma_{t+n})]$$

واختزلنا :

$$\xi_{t+1} - \xi_t = \frac{1}{2n+1} (\gamma_{t+n+1} - \gamma_{t-n}).$$

ولكن من المفترض أن تكون γ_t دالّة دورية ، دورتها 2n+1 :

$$\gamma_{t+n+1} = \gamma_{t-n}.$$

إذاً :

$$\xi_{t+1} - \xi_t = 0$$

وبالتالي ξ_t هو كمّية ثابتة .

إذا افترضنا بالإضافة إلى هذا أنّ حاصل جمع المكوّنات الموسمية على 2n+1 دورة يساوي صفراً :

$$\sum_{k=-n}^{+n} \gamma_{t+k} = 0$$

يصبح المتوسّط المتحرّك لهذه المكوّنات يساوي صفراً . عندئذٍ يقتصر متوسّط السلسلة الزمنية المتحرّك على المتوسّط المتحرّك للمكوّنة الظرفية :

$$z_t = M_{2n+1}(G_t)$$

إنّ العملية M_{2n+1} توقف كلّياً الدالاّت الدورية ذات الدورة 2n+1 .

بالتالي ، إذا أجرينا ، في تحليل سلسلة زمنية تتضمّن تغيّرات موسمية معروفة الدورة (مثلاً 12 شهراً) ، عمليّة « المتوسّط المتحرّك » بطول يساوي هذه الدورة ، فإنّ هذه العملية « تحذف » المكوّنة الموسمية إذا كانت دورية تماماً ، هذه الخاصة هي وراء طرق تجريبية عدّة لتصحيح التغيّرات الموسمية . يجب أيضاً التأكّد من أنّ هذه العملية لا « تحرّف » الاتجاه غير الموسمي ولا تدخل تطوّرات غير قابلة للتفسير .

B . تصفية المكوّنة غير الموسمية

أ ـ الاتجاه غير الموسمي الخطّي

لنفترض أنّ المكوّنة غير الموسمية هي دالّة خطّية تبعاً للوقت :

$$c_t = at + b .$$

ولنحسب متوسّطها المتحرّك ، بطول $p = 2n + 1$:

$$z_t = \frac{1}{2n+1}\sum_{k=-n}^{n} c_{t+k} = \frac{1}{2n+1}\sum_{k=-n}^{n}[a(t+k)+b]$$
$$= \frac{a}{2n+1}\sum_{k=-n}^{n}(t+k)+b .$$

إذا وسّعنا ، يصبح لدينا ، بحكم تناظر السلسلة بين الرمزين [ ] :

$$z_t = \frac{a}{2n+1}[(t-n)+(t-n+1)+\cdots+(t-1)+t+(t+1)+\cdots+(t+n-1)+(t+n)]+b$$
$$= \frac{a}{2n+1}(2n+1)t+b = at+b .$$

بالتالي ، إذا حلّلنا على طريقة المتوسّطات المتحرّكة ، سلسلة زمنية تكون مكوّنتها غير الموسمية خطّية ، فإنّ هذه المكوّنة تمرّ في المصفاة دون أن تتأثّر .

ب ـ أيّ اتجاه غير موسمي

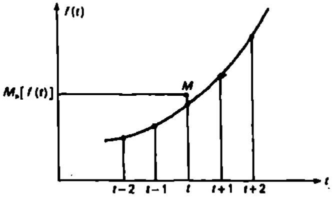
لنفترض أنّ المكوّنة غير الموسمية هي دالّة تبعاً للوقت :

$$c_t = f(t) .$$

إنّ المتوسّط المتحرّك ، بطول 2n+1 ، المنقول في التاريخ t يطابق مركز الثقل M للـ 2n+1 نقطة التي تحيط بهذا التاريخ (الشكل 62) .

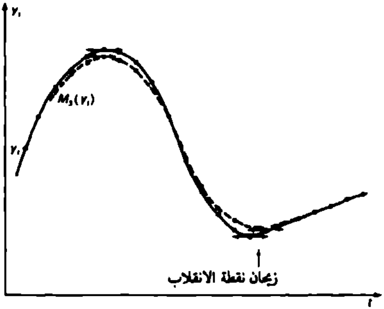
بما أنّ مركز الثقل يقع حتماً في تجويف المنحنى :

ـ يكون المتوسّط المتحرّك أكبر من f(t) إذا كان تحويف المنحنى نحو الأسفل ؛
ـ إذا كانت f(t) دالّة خطّية ، فمركز الثقل يوجد على المنحنى : إنّ عملية « المتوسّط المتحرّك » تحوّل الخطّ إلى نفسه ، كما أبرزنا في الفقرة السابقة .



الشكل 62

عندما تتضمّن السلسلة الزمنية نقاط انقلاب في الإتّجاه ، فالسلسلة المشتقّة بواسطة عملية « المتوسّط المتحرّك » تتضمّن ، هي أيضاً ، نقاط انقلاب ، ولكن ، إذا كان المنحنى غير متناظر ، قد تكون هذه النقاط مزاحة ، إلى الأمام أو إلى الخلف حسب الحالة (الشكل 63) . وهذا الاحتمال مزعج ، في الواقع ، عندما نعمد إلى تحليل سلسلة زمنية ، نحاول بشكل عام أن نتكهّن ، أو على الأقلّ أن نستنتج بأسرع ما يمكن « انقلابات » الظرف . إذا أدّت طريقة التحليل إلى إزاحة هذه النقاط ، هنا إمكانية كبيرة للوقوع في الخطأ .



الشكل 63 .
وضع السلسلة الزمنية ومتوسّطها المتحرّك

لنأخذ بأنّ المتوسّط المتحرّك يحوّل اتّجاهاً غير موسمي إلى اتجاه قريب بما يكفي : يكون تقريب الاتجاه الحقيقي أفضل كلّما اقترب هذا الاتجاه من خطّ مستقيم . عند وجود نقاط انقلاب ، لا تكون دقّة الطريقة كاملة .

C . تصفية التقلّبات العشوائية

لندرس الآن تأثير عملية « المتوسّط المتحرّك » على سلسلة البواقي العشوائية ε_t التي تدخل في تجزئة السلسلة الزمنية التالية :

$$y_t = c_t + s_t + \varepsilon_t .$$

لنفترض أنّ

$$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_t, \ldots, \varepsilon_T$$

هي متتالية من المتغيّرات العشوائية المستقلّة بأمل رياضي يساوي صفراً وبانحراف نموذجي ثابت :

$$E(\varepsilon_t) = 0 \quad , \qquad V(\varepsilon_t) = \sigma^2, \quad \forall t .$$

لنرمز بواسطة η_t إلى المتوسّط المتحرّك بطول 2n+1 لسلسلة التقلّبات العشوائية هذه :

$$\eta_t = \frac{1}{2n+1} \sum_{k=-n}^{+n} \varepsilon_{t+k} , \quad (t = n+1, n+2, \ldots, T-n)$$

وبفضل خصائص الأمل الرياضي والتباين (الفصل I ، ص 55 و61) .

$$E(\eta_t) = \frac{1}{2n+1} \sum_{k=-n}^{+n} E(\varepsilon_{t+k}) = 0$$ لأنّ $E(\varepsilon_{t+k}) = 0$ مهما تكن k ؛

$$V(\eta_t) = \frac{1}{(2n+1)^2} \sum_{k=-n}^{+n} V(\varepsilon_{t+k}) = \frac{1}{(2n+1)^2} \sum_{k=-n}^{+n} \sigma^2 = \frac{\sigma^2}{2n+1}$$

بعد المرور في « المصفاة » ، يصبح تباين η_t أصغر بـ (2n+1) مرّة من تباين السلسلة الأصلية : إذن نخفّف التقلّبات العشوائية إلى حدّ بعيد ، ونقول أنّ المتوسّط المتحرّك « يصقل » السلسلة .

إلّا أنّ لهذا الإجراء ناحية سلبية وهي أنّ المتغيّرات η_t التي تنتج عنه لا تعود مستقلّة كما الحال مع التقلّبات ε_t الأصلية . ومتغيّرتان η متجاورتان :

$$\eta_t = \frac{1}{2n+1} \sum_{k=-n}^{n} \xi_{t+k}$$

و

$$\eta_{t+1} = \frac{1}{2n+1} \sum_{k=-n}^{n} \xi_{t+1+k}$$

تشتركان في 2n−2 متغيّرة ξ_i :

$$\xi_{t-n+1}, \dots, \xi_{t+n}$$

وتكونان على ارتباط وثيق . وهـذا الارتباط قد يولّد ، خاصّة إذا كنّا نكرّر عمليـة « المتوسّـط المتحرّك » كما سنـرى لاحقاً ، حـركات دوريـة لم تكن موجـودة في السلسلة الأصليـة . ولقد حـذّر عالم الإحصـاء الـروسي سلوتسكي Slutsky من هـذه الظاهرة .

بالمختصر ، يسمح تطبيق « المتوسّـط المتحرّك » على سلسلة زمنية :

- بحذف المكوّنة الموسمية إذا كانت دورية تماماً ؛
- بالاحتفاظ تقريباً بالمكوّنة غير الموسمية طالما لم يكن انحناؤها قويّاً ؛
- بصقل التقلّبات المتبقّية .

على هذه المجموعة من الخصائص يستند تصحيح التغيّرات الموسمية على طريقة المتوسّطات المتحرّكة . وكي يمكن إجراء هذا التصحيح ، يجب ملء عدد من الشروط المتعلّقة بالعناصر التي تكوّن السلسلة الزمنية .

3 . تصحيح التغيّرات الموسمية

A . الفرضيات

نسلّم بأنّ السلسلة الزمنية مؤلّفة من ثلاث مكوّنات : غير الموسمية ، الموسمية والمتبقّية .

قد تكون صورة تكوين هذه العناصر (أنظر سابقاً ، الفقرة 2 ، ص 364) :

- إمّا جمعية :

$$y_t = c_t + s_t + \varepsilon_t$$

- إمّا مضاعفة :

$$y_t = c_t(1 + s_t) + \varepsilon_t$$

(حيث المكوّنة الموسمية $c_t s_t$ تناسبية مع الحركة الظرفية) أو :

$$y_t = c_t(1 + s_t)(1 + \varepsilon_t)$$

(حيث المكوّنة الموسمية تناسبية مع الحركة الظرفية ،والتغيّر المتبقّي ε_t $c_t(1+s_t)$ تناسبي مع مجموع المكوّنتين الأوليين) .

إذا أخذنا لوغاريتم العنصرين ، يتحوّل هذا الشكل الثاني إلى صورة جمعية .

أ ـ الفرضيات المتعلّقة بالحركة غير الموسمية

الحركة غير الموسمية c_t هي دالّة تبعاً للوقت ، لا تتضمّن انقلاباً ذا اتجاه لافت أو ملحوظ جدّاً .

ضمن هذه الشروط ، يمكننا أن نقرّ بأنّ عملية « المتوسّط المتحرّك » تحوّل c_t إلى دالّة قريبة جدّاً :

$$M_p(c_t) \doteq c_t . \qquad (1)$$

ب ـ الفرضيات المتعلّقة بالتغيّرات الموسمية

نفترض أنّ :

ـ المكوّنة الموسمية s_t هي دالّة دورية تماماً ، ودورتها p ، تأخذ القيم s_j (j=1, 2, .., p) :

$$s_t = s_{tj} = s_j , \qquad s_{t+p} = s_{t+1,j} = s_j ;$$

ـ مجموع الـ p مكوّنة موسمية s_j يساوي ، بناء على التعريف ، صفراً :

$$\sum_{j=1}^{p} s_j = 0 ,$$

بشكل تعوّض فيه ، في الدورة الواحدة ، المكوّنات الموسمية الإيجابية تماماً عن المكوّنات الموسمية السلبية .

ضمن هذه الشروط ، يكون المتوسّط المتحرّك بطول p ، المطبّق على s_t ، يساوي صفراً :

$$M_p(s_t) = 0 . \qquad (2)$$

ج ـ الفرضيات المتعلّقة بالتقلّبات المتبقّية

نقرّ بأنّ التقلّبات المتبقّية ε_t هي متغيّرات عشوائية مستقلّة عن الحركة الظرفية

وعن التغيّرات الموسمية ، وأنّ أملها الرياضي يساوي صفراً وتباينها ضعيف :

$$E\{\varepsilon_t\} = 0, \qquad V\{\varepsilon_t\} \neq 0.$$

بالتالي ، يتضمّن المتوسّط المتحرّك للتغيّرات المتبقّية تقلّبات أضعف أيضاً حول الصفر :

$$M_p(\varepsilon_t) \neq 0. \qquad (3)$$

إلاّ أنّ التمعّن في مكوّنات سلسلة زمنية معيّنة قد أظهر أنّه يمكننا تصنيف التغيّرات المتبقّية في فئتين . يمكننا إرجاع العدد الأكبر منها إلى كمّية كبيرة من الأسباب الصغيرة، أخطاء القياس بصورة خاصّة، التي تُحدِث في الواقع تغيّرات ضعيفة المدى. ولكن البعض الآخر ينتج عن حوادث عرضية منفصلة وواسعة المدى : إضراب ، قرار إداري ، انهيار مالي ، كارثة طبيعية ، الخ .

وحده النوع الأوّل من الاحتمالات يلبّي الفرضية المطروحة . إذن كي يمكننا تطبيق الطريقة ، يصبح من الضروري أن نصحّح مسبقاً المعطيات الملحوظة الخام المتعلّقة بالتغيّرات العرضية المهمّة . بشكل عام ، يكون من السهل أن نعاين على رسومات بيانية منفّذة[1] ، المعطيات التي تتضمّن تقلّبات كبيرة . وعندئذٍ نصحّحها إمّا بتقدير أثر الظاهرة العرضية مباشرة ، إمّا بواسطة تقييم بياني بسيط كما هو الحال معظم الأحيان .

ضمن الفرضيات السابقة ، يحوّل المتوسّط المتحرّك بطول p سلسلة المعطيات الملحوظة إلى سلسلة قريبة من الحركة غير الموسمية .

في حالة صورة جمعية :

$$y_t = c_t + s_t + \varepsilon_t,$$

نأخذ هذه النتيجة مباشرة من إحدى خصائص المتوسّط المتحرّك كمؤثّر خطّي ومن العلاقات (1) ،(2) و(3) .

$$M_p(y_t) = M_p(c_t) + M_p(s_t) + M_p(\varepsilon_t) \neq c_t.$$

---

(1) انظر لاحقاً ، ص 386 .

وفي حالة صورة مضاعِفة :

$$y_t = c_t(1 + s_t) + \varepsilon_t = c_t + c_t s_t + \varepsilon_t ,$$

يجب ، بالإضافة إلى هذا ، أن نفترض أنّ الحركة غير الموسمية لا تتغيّر كثيراً في الدورة الواحدة ويمكن اعتبارها بالتالي ثابتة ومساوية لمتوسّطها تقريباً :

$$c_t \doteq \bar{c} .$$

بالتالي :

$$c_t s_t \doteq \bar{c} s_t$$

$$M_p(c_t s_t) \doteq \bar{c} M_p(s_t) = 0$$

و

$$M_p(y_t) = M_p(c_t) + M_p(c_t s_t) + M_p(\varepsilon_t) \doteq c_t .$$

B . حساب المعامِلات الموسمية

نعرض في ما يلي مختلف مراحل تحليل السلسلة الزمنية ، ونوضّحها في الفقرة اللاحقة بواسطة مثل تطبيقي .

1 . وضع الرسومات البيانية المنضّدة

يسمح رسم المنحنيات المنضّدة بيانياً (الشكل 65 ، ص 386) بإبراز وجود تغيّرات موسمية دورتها p وباختيار صورة التكوين المناسبة .

يمكن رسم المنحنيات المنضّدة على بيانات ذات إحداثيات حسابية أو نصف لوغاريتمية .

على رسم بياني حسابي ، إذا كان مدى الحركة الموسمية تقريباً ثابتاً فهذا يشير إلى صورة جمعية : الحركة الموسمية مستقلّة عن المستوى الذي تصل إليه السلسلة . أمّا إذا كان المدى (أو الذروة) يتغيّر مع مستوى السلسلة فهذا يوحي بصورة مضاعِفة .

على رسم بياني نصف لوغاريتمي ، إذا كان مدى الحركة الموسمية تقريباً ثابتاً فهذا يشير إلى صورة مضاعفة : الحركة الموسمية هي تناسبية مع المستوى الذي تصل اليه السلسلة .

2 . تصحيح التغيّرات العرضية كبيرة المدى

بشكل عام ، يكفي التمعّن في الرسومات البيانية المنضّدة للانتباه إلى الشواذات

المحتملة في تطوّر الكمّية موضع الدراسة . ومن الضروري أن نحيط علماً بشكل كامل بالدورة المدروسة كي نكشف سبب هذه التغيّرات العرضية ذات المدى الإستثنائي (إضراب ، حادث مناخي ، الخ) .

ويتمّ تصحيح المعطيات الخام « غير الطبيعية » إمّا بواسطة تقدير مباشر (مثلاً ، تقييم الخسارة في الانتاج التي يحدثها إضراب) ، إمّا بواسطة تقدير بياني بسيط .

3 . حساب المتوسّط المتحرّك

بشكل عام ، تكون دورة التغيّرات الموسمية p مزدوجة (12 شهراً أو 4 فصول مثلاً) . يتمّ حساب المتوسّط المتحرّك بطول p ، والمتعلّق بالتاريخ t ، حسب الاصطلاح المعروض أعلاه (ص 369) . هكذا يصبح المتوسّط المتحرّك على 12 شهراً :

$$M_{12}(y_t) = \frac{1}{12}\left[\frac{1}{2}y_{t-6} + (y_{t-5} + y_{t-4} + \cdots + y_{t+5}) + \frac{1}{2}y_{t+6}\right]$$

ويتنوّع ما تبقّى من الحساب تبعاً لما نفترض ، صورة جمعية أم مضاعِفة .

الصورة الجمعية

نكتب الصورة الجمعية :

$$y_t = c_t + s_t + a_t$$

بحكم الفرضيات المطروحة :

$$y_{ij} = c_{ij} + s_j + a_{ij} ;$$

ونسمّي s_j « المعامِل الموسمي » .

4 . حساب الفوارق مع المتوسّط المتحرّك $\quad d_{ij} = y_{ij} - M(y_{ij})$.

5 . تركيب الفوارق الموسمية

بالنسبة لكلّ « شهر » j ، نحدّد تقديراً أوّل s'_j للمعامِل الموسمي بأخذنا وسيط الفوارق الموسمية المتعلّقة بهذا الشهر ، أو متوسّطها بعد حذفنا احتمالياً الفوارق الشاذّة[1] .

(1) بشكل عام ، نفضّل أخذ الوسيط عن أخذ المتوسّط الحسابي ، من أجل تخفيف الأثر المحتمل للفوارق الموسمية غير الطبيعية . عندما يكون الحساب على حاسب آلي ، نعتمد غالباً المتوسّط ، ولكن بعد إبعاد الفوارق الطرفية التي قد تكون شاذّة .

في الواقع ، حيث :

$$M(y_{ij}) \# c_{ij} .$$

يكون لدينا :

$$d_{ij} \# s_j + \varepsilon_{ij}$$

وإذا أخذنا الأمل الرياضي :

$$E\{d_{ij}\} \# s_j$$

لأنّ $E\{\varepsilon_{ij}\} = 0$ بحكم الفرضيات المطروحة .

بالتالي ، إذا أخذنا متوسّط الفوارق d_{ij} أو وسيطها ، نحصل على تقدير المعامل الموسمي s_j بواسطة $\hat{s}_j$.

6 . التقدير النهائي للمعاملات الموسمية

بناء على التعريف ، يجب أن تحقّق المعاملات الموسمية العلاقة التالية :

$$\sum_{j=1}^{p} s_j = 0 .$$

بما أنّ التقديرات $\hat{s}_j$ جرت كلًّا على حدة انطلاقاً من سلاسل الفوارق الموسمية المتعلّقة بكلّ « شهر » ، فإنّ مجموعها لا يكون بشكل عام مطابقاً للصفر . إذاً ، نحصل على التقديرات النهائية s_j^* للمعاملات الموسمية بتصحيحنا التقديرات الأولى بشكل يراعي علاقة التعريف هذه :

ـ حساب متوسّط الـ p تقدير $\hat{s}_j$:

$$\bar{s} = \frac{1}{p}\sum_{j=1}^{p} \hat{s}_j ;$$

ـ تصحيح المعاملات الموسمية :

$$s_j^* = \hat{s}_j - \bar{s} .$$

يمكن كذلك إجراء تصحيح التقديرات الأولى للمعاملات الموسمية مع الأخذ بعين الاعتبار نسبة الشك في كلّ منها التي نقيسها بواسطة الانحراف النموذجي للفوارق الموسمية المتعلّقة « بالشهر » المناسب . ويكون انحراف مجموع المعاملات $\hat{s}_j$ عن الصفر

موزّعاً ، بهذه الطريقة ، تناسبياً مع الانحرافات النموذجية للفوارق الموسمية المتعلّقة بكلّ « شهر » والمحسوبة بعد إبعاد محتمل للفوارق الشاذّة .

7 . وضع السلسلة مصحّحة التغيّرات الموسمية y_{ij}^{*}

$$y_{ij}^{*} = y_{ij} - s_j^{*} .$$

إذا كان تقدير المعايلات الموسمية صحيحاً ، فإنّ y_{ij}^{*} تساوي حاصل جمع المكوّنة الظرفية c_{ij} مع التغيّر العشوائي ε_{ij} ،وكوننا فترضنا أنّ مدى هذا الأخير ضعيف ، فإنّ السلسلة مصحّحة التغيّرات الموسمية تمثّل تقريباً جيّداً لاتجاهات تطوّر الكمّية الملحوظة .

الصورة المضاعفة

إنّ الصورة المضاعفة :

$$y_t = c_t(1 + s_t) + \varepsilon_t$$

تُكتب بفضل الفرضيات المطروحة :

$$y_{ij} = c_{ij}(1 + s_j) + \varepsilon_{ij} .$$

وإذا وضعنا :

$$S_j = 1 + s_j$$

نحصل على :

$$y_{ij} = c_{ij}\, S_j + \varepsilon_{ij} .$$

نسمّي S_j « المعامل الموسمي » .

4 . حساب النسب على المتوسّط المتحرّك :

$$r_{ij} = \frac{y_{ij}}{M(y_{ij})} .$$

5 . تركيب النسب الموسمية

بالنسبة لكلّ « شهر » j ، نقوم بتقدير أوّل S_j للمعامل الموسمي بأخذنا وسيط النسب الموسمية المتعلّقة بهذا الشهر ، أو متوسّطها بعد حذف محتمل للنسب الشاذّة[1] .

---

(1) أنظر الملاحظة السابقة .

في الواقع وحيث :

$$M(y_{ij}) \simeq c_{ij} \,.$$

يكون لدينا :

$$r_{ij} \simeq S_j + \frac{\varepsilon_{ij}}{M(y_{ij})}$$

وإذا أخذنا الأمل الرياضي :

$$E\,(r_{ij}) \simeq S_j + E\left\{\frac{\varepsilon_{ij}}{M(y_{ij})}\right\} .$$

لكن بحكم الفرضيات المطروحة حول التغيّرات المتبقّية وبما أنّه يمكننا اعتبار ε_{ij} و $1/M(y_{ij})$ عمليّاً مستقلّين ، لدينا :

$$E\left\{\frac{\varepsilon_{ij}}{M(y_{ij})}\right\} \simeq 0 .$$

إذن ، إذا أخذنا متوسّط النسب الموسمية r_{ij} أو وسيطها نحصل على تقدير للمعامل الموسمي S_j وهو $\hat{S}_j$.

6 . التقدير النهائي للمعاملات الموسمية

بناء على تعريف المكوّنة الموسمية :

$$\sum_{j=1}^{p} s_j = 0$$

إذن :

$$\sum_{j=1}^{p} S_j = \sum_{j=1}^{p} (1 + s_j) = p .$$

أي أنّ مجموع المعاملات الموسمية S_j يساوي p .

بما أنّنا قمنا بالتقديرات $\hat{S}_j$ كلّ على حدة انطلاقاً من سلاسل النسب الموسمية المتعلّقة بكلّ « شهر » ، فإنّ مجموعها لا يساوي p بشكل عام . إذن نحصل على التقديرات النهائية S_j^* للمعاملات الموسمية بتصحيحنا تناسبياً التقديرات الأولى بشكل يراعي هذه العلاقة :

ـ حساب متوسّط الـ p تقدير $\hat{S}_j$.

$$\bar{S}' = \frac{1}{p} \sum_{j=1}^{p} S_j' ,$$

ـ تصحيح المعاملات الموسمية :

$$S_j^* = \frac{S_j'}{\bar{S}'} .$$

يمكن كذلك إجراء تصحيح التقديرات الأولى للمعاملات الموسمية مع الأخذ بعين الاعتبار نسبة الشك في كلّ منها التي نقيسها بواسطة الانحراف النموذجي للنسب الموسمية المتعلّقة « بالشهر » المناسب . ويكون الانحراف بين مجموع المعاملات S_j' وعدد « الأشهر » التي تؤلف الدورة ، بهذه الطريقة ، موزّعاً تناسبياً مع الانحرافات النموذجية للنسب الموسمية المتعلّقة بكلّ « شهر » والمحسوبة بعد إبعاد محتمل للنسب الشاذة .

4 . وضع السلسلة مصحّحة المتغيّرات الموسمية y_{ij}^*

$$y_{ij}^* = \frac{y_{ij}}{S_j^*} .$$

C . مثل تطبيقي : المؤشّر الفصلي للانتاج الصناعي

إنّ المؤشّر الفصلي للانتاج الصناعي (دون البناء والأشغال العامّة) ، بقاعدة 100 في العام 1962 ، يكامل بين عدد من المعطيات بتردّد فصلي ، لا تظهر إذن في المؤشّر الفصلي (منشآت الطيران ، صناعة الآلات والأجهزة الميكانيكية ، الصناعات الزراعية والغذائية ، الخ) .

نعرض سلسلة المؤشّرات الفصلية للسنوات من 1962 إلى 1969 في الجدول 32 . أمّا تمثيلها البياني (الشكل 64) ، الذي يُظهر تغيّرات موسمية مهمّة ، لا يسمح ، كما هو ، بتحليلٍ مرضٍ لاتجاهاتٍ تطوّر الانتاج الصناعي . هنا يبدو تصحيح التغيّرات الموسمية ضرورياً .

1. الرسم البياني للمنحنيات المنضّدة

بالإحداثيات الحسابية (الشكل 65) ، تظهر المنحنيات المنضّدة حركة موسمية يتزايد مداها مع مستوى السلسلة . بالمقابل ، على رسم بياني نصف لوغاريتمي (الشكل 66) ، يظهر مدى الحركة الموسمية تقريباً ثابتاً : علينا إذن أن نعتمد صورة مضاعفة .

الجدول 32 . المؤشّر الفصلي للانتاج الصناعي
(ما عدا البناء والأشغال العامّة)

المصدر : INSEE (القراءة من اليسار إلى اليمين)

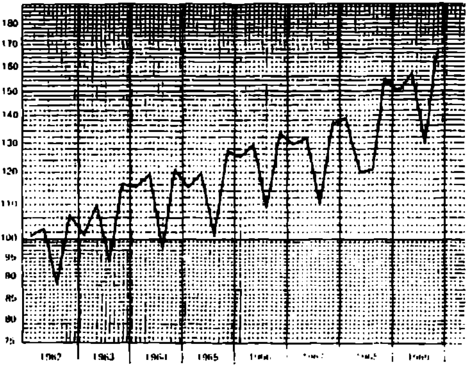
السنة	الفصل الأوّل	الفصل الثاني	الفصل الثالث	الفصل الرابع
1962	101,3	102,9	88,4	107,3
1963	101.0 ([1])	109,8	94,1	116,1
1964	115.6	119,2	97,7	120,3
1965	115,1	119,5	101,1	127,4
1966	124,8	129,0	109,3	133,6
1967	129,4	131,8	110,2	136,4
1968	138,5	120,1 ([2])	120,8	154,4
1969	149,5	157,1	130,8	166,5

تصحيح التغيّرات العرضية الإستثنائية :

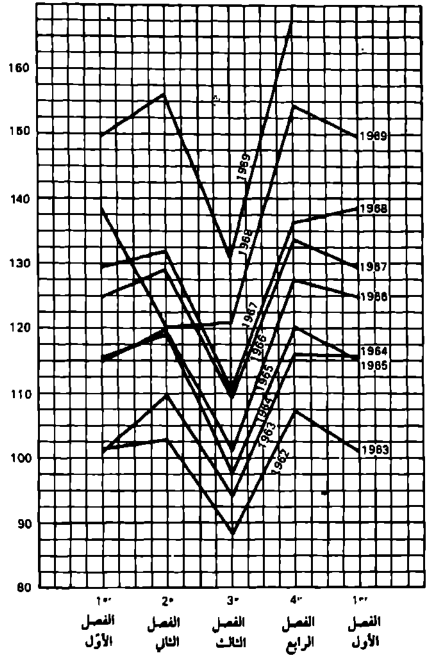
(1) شتاء قارس بشكل استثنائي وإضراب عمّال المناجم في آذار 1963 . التصحيح المقترح : 107,5

(2) الاضراب العام في أيار ـ حزيران 1968 . التصحيح المقترح : 141,0 .

الشكل 64 . المؤشّر الفصلي للانتاج الصناعي ، القاعدة 100 في عام 1962 . المعطيات الخام . الإحداثيات الصادية لوغاريتمية



الشكل 65 . المؤشّر الفصلي للانتاج الصناعي .
المنحنيات المتنضّدة . الإحداثيات حسابية .

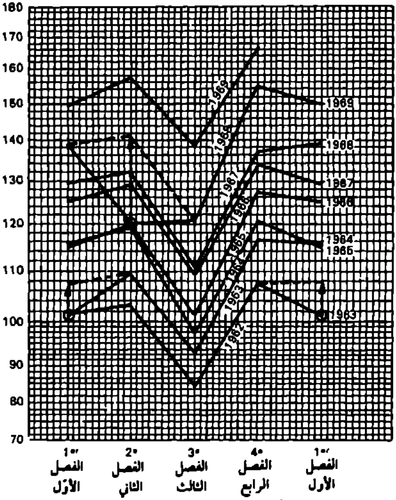


2 . تصحيح التغيّرات العرضية الإستثنائية

نرى على الرسم البياني للمنحنيات المتنضّدة (الشكل 66) وبوضوح شواذين : إنهما يتعلّقان بالفصل الأوّل من العام 1963 والفصل الثاني من العام 1968 . في الواقع ، يتطابق الشواذ الأوّل مع قساوة الطقس الإستثنائية في شتاء 1962 -1963 وإضراب عمّال المناجم في آذار 1963 ، والشواذ الثاني مع الإضراب العام في أيّار -

حزيران 1968 . من أجل حساب المتوسّط المتحرّك والنسب الموسمية ، تمّ تصحيح هاتين المعطيتين غير الطبيعيتين :

	المعطية الخام	المعطية المصحّحة
الفصل الأوّل 1963	101,0	107,5
الفصل الثاني 1968	120,1	141,0



شكل 66 المؤشّر الفصلي للانتاج الصناعي . المنحنيات المتقطّعة.
الإحداثيات الصادية لوغاريتمية.

3 . حساب المتوسّط المتحرّك

لقد تمّ حساب المتوسّط المتحرّك على 4 فصول بواسطة الحاسب الآلي . ونعرض النتائج في الجدول 33 .

الحساب اليدوي يتمّ بالطريقة التالية :

ـ حساب المجموعات المتحرّكة المنقولة عند منتصف الدورة :

$$s(t-\tfrac{1}{2}) = \sum_{k=-2}^{+1} y_{t+k}.$$

ننتقل من مجموع متحرّك إلى تابعه بطرحنا المشاهدة الأولى وبإضافتنا المشاهدة المناسبة في السنة التي تلي :

$$s(t+\tfrac{1}{2}) = s(t-\tfrac{1}{2}) - y_{t-2} + y_{t+2}.$$

ـ حساب حواصل جمع مجموعين متحرّكين متتاليين :

$$S(t) = s(t-\tfrac{1}{2}) + s(t+\tfrac{1}{2}).$$

الجدول 33 . المؤشّر الفصلي للانتاج الصناعي
(ما عدا البناء والأشغال العامّة) .

I . المتوسّطات المتحرّكة على 4 فصول
(القراءة من اليسار إلى اليمين)

السنة i	الفصل j	1	2	3	4
1	1962	—	—	100,8	102,4
2	1963	104,0	105,8	107,9	110,1
3	1964	111,7	112,7	113,1	113,1
4	1965	113,6	114,9	117,0	119,4
5	1966	121,6	123,4	124,8	125,7
6	1967	126,1	126,6	128,1	130,4
7	1968	132,9	136,4	140,1	143,4
8	1969	146,7	149,5	—	—

ـ حساب المتوسّطات المتحرّكة :

$$M_4(t) = \frac{1}{8}\left[s\left(t-\frac{1}{2}\right) + s\left(t+\frac{1}{2}\right)\right] = \frac{S(t)}{8}.$$

II . حساب النسب الموسمية وتقدير المعاملات الموسمية

السنة i	الفصل j	1	2	3	4
1	1962	—	—	0,877 4	1,048 0
2	1963	1,034 0	1,038 1	0,872 2	1,054 7
3	1964	1,034 9	1,057 9	0,863 6	1,063 5
4	1965	1,013 4	1,040 1	0,864 2	1,067 1
5	1966	1,026 3	1,045 4	0,876 2	1,063 1
6	1967	1,025 9	1,041 1	0,860 3	1,046 2
7	1968	1,042 5	1,033 5	0,862 5	1,076 4
8	1969	1,019 1	1,051 1	—	—
المعاملات الموسمية	التقدير الأوّل S'_j	1,028 0	1,043 1	0,867 7	1,059 3
	التقدير النهائي S_j^*	1,028 5	1,043 6	0,868 1	1,059 8

مثلاً : تطبيق هذه الحسابات على بداية السلسلة :

القيم الخام[1] y_t	حواصل الجمع المتحركة $s(t-1/2)$	حواصل جمع متتاليين مجموعين متحركين $S(t)$	المتوسّطات المتحركة $M_4(t)$
101,3	—	—	—
102,9	—	—	—
	399,9		
88,4		806,0	100,8
	406,1		
107,3		819,1	102,4
	413,0		
107,5		831,7	104,0
	418,7		
109,8		846,2	105,8
	427,5		
94,1	⋮	⋮	⋮
116,1			
⋮			

(1) بعد تصحيح التغيّرات العرضية الإستثنائية .

4 . حساب النسب على المتوسّط المتحرّك

نعرض نتائج هذا الحساب :

$$r_{ij} = \frac{y_{ij}}{M_4(y_{ij})}$$

في الجدول 33 .

5 . تركيب النسب الموسمية

لقد تمّ تركيب النسب الموسمية على الحاسب الآلي بأخذ متوسّطها ، بعد استبعاد أكبرها وأصغرها في الواقع ، يُحتمل أن تكون النسبتان الطرفيتان قيمتين شاذّتين . وتظهر هذه التقديرات الأولى S'_j للمعامِلات الموسمية عند أسفل الجدول 33 .

من الضروري إجراء فحص مدروس لقيمة المعاملات الناتجة عن هذا الإجراء الآلي . وقد تمّ هذا الأمر على رسوم بيانية من النوع المعروض في الشكل 67 . في الإجراء الآلي ، توضع هذه الرسوم بواسطة الحاسب . وعلى هذه الرسوم ، تظهر القيم الطرفية ، المستبعدة عن حساب المعامل الموسمي ، محاطة بدوائر صغيرة وتظهر القيمة المقدّرة للمعامل الموسمي ممثّلة بخطّ أفقي منقّط . نستنتج أنّ إعتماد الوسيط للقيام بتركيب النسب الموسمية يعطي قيم معامِلات موسمية مختلفة بعض الشيء وأقلّ ملاءمة .

6 تقدير المعاملات الموسمية نهائياً

يجب أن يكون مجموع المعاملات الموسمية 4 (عدد فصول السنة) :

$$\sum_{j=1}^{4} S_j = \sum_{j=1}^{4} (1 + s_j) = 4 ,$$

لأنّه ، بناء على تعريف المكوِّنة الموسمية :

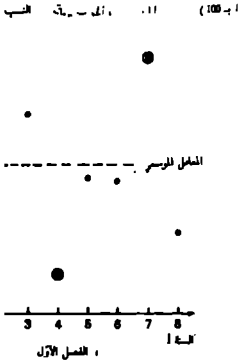
$$\sum_{j=1}^{4} s_j = 0$$

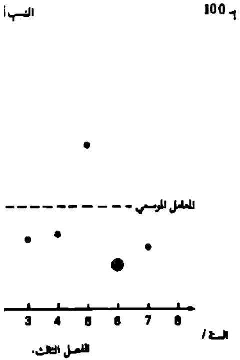
ولكن في الحقيقة لا يتطابق مجموع التقديرات الأولى S'_j للمعامِلات الموسمية مع 4 :

$$\sum_{j=1}^{4} S'_j = 3{,}998\ 1 .$$

نحسب التقديرات النهائية S_j^* للمعاملات الموسمية بتصحيحنا تناسبياً التقديرات الأولى S'_j :

$$S_1^* = \frac{1{,}028\ 0 \times 4}{3{,}998\ 1} = 1{,}028\ 5$$


المعامل الموسمي
3 4 5 6 7 8
الفصل الأول


المعامل الموسمي
3 4 5 6 7 8
الفصل الثالث

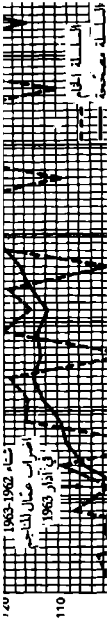
$$S_2^* = \frac{1,043\,1 \times 4}{3,998\,1} = 1,043\,6$$

$$S_3^* = \frac{0,867\,7 \times 4}{3,998\,1} = 0,868\,1$$

$$S_4^* = \frac{1,059\,3 \times 4}{3,998\,1} = 1,059\,8\,.$$

يمكننا إجراء التقدير النهائي للمعاملات الموسمية بطريقة منطقية أكثر بتوزيعنا الانحراف بين مجموع المعاملات S_i^* والعدد 4 ، تناسبياً مع الانحرافات النموذجية للنسب الموسمية المتعلّقة بكل فصل ، والمحسوبة بعد استبعاد أصغرها وأكبرها . ميزة هذه الطريقة أنّها تأخذ بعين الاعتبار نسبة الشكّ الفعلي المتعلّقة بكلّ من هذه التقديرات . في هذا المثل ، النتائج الحاصلة مختلفة قليلاً جدّاً عن النتائج التي أعطتنا إيّاها الطريقة الأولى :

	الفصل الأوّل	الفصل الثاني	الفصل الثالث	الفصل الرابع
التباين المصحّح	0,000034	0,000022	0,000030	0,000049
الانحراف النموذجي المصحّح	0,006	0,005	0,005	0,007
المعامل الموسمي	1,0285	1,0435	0,8681	1,0599


1963-1962
1963
110

7 . وضع السلسلة مصحّحة التغيّرات الموسمية

• نحصل على المعطيات مصحّحة التغيّرات الموسمية بقسمتنا المعطيات الخام ، قبل تصحيح التغيّرات العرضية في الفصل الأوّل عام 1963 والفصل الثاني 1968 ، على المعامل الموسمي للفصل المناسب :

$$y^*_{ij} = \frac{y_{ij}}{S^*_j}.$$

نعرض نتائج هذه الحسابات في الجدول 34 ، وقد مثّلنا السلسلة مصحّحة التغيّرات الموسمية على ذات الرسم البياني نصف اللوغاريتمي للسلسلة الخام (الشكل 68) . فيما لم يكن بالإمكان إعطاء أيّ حكم دقيق بالنسبة للسلسلة الخام ، فإنّ السلسلة مصحّحة التغيّرات الموسمية تسمح لنا أن نتابع تطوّر الانتاج الصناعي فصلاً فصلاً : بعد التزايد السريع في العام 1963 بنسبة سنوية مقدارها 10% ، نلاحظ نوعاً من الركود عند نهاية العام 1964 ، ثمّ تزايداً معتدلاً بنسبة قريبة من 6% في السنة خلال العامين 1965 و1966 ، وبطءاً عند نهاية العام 1966 وخلال العام 1967 ، وأخيراً تسارعاً بعد شهر آيار 1968 بنسبة تزايد سنوية مقدارها 9% . ويسمح لنا تمثيل السلسلة على ورق نصف لوغاريتمي بتقييم بياني مباشر لنسب التزايد .

الجدول 34 . المؤشّر الفصلي للانتاج الصناعي
(ما عدا البناء و الأشغال العامّة)

السلسلة مصحّحة التغيّرات الموسمية

(القراءة من اليسار إلى اليمين)

السنة	الفصل الأوّل	الفصل الثاني	الفصل الثالث	الفصل الرابع
1962	98,5	98,6	101,8	101,2
1963	98,2	105,2	108,4	109,5
1964	112,4	114,2	112,5	113,5
1965	111,9	114,5	116,5	120,2
1966	121,3	123,6	125,9	126,1
1967	125,8	126,3	126,9	128,7
1968	134,7	115,1	139,2	145,7
1969	145,4	150,5	150,7	157,1

الملحقات

جدول 1 . قانون بواسّون (Poisson)

الترددات الفردية : $P_x = \frac{e^{-m} m^x}{x!}$.

الترددات المتراكمه $F(x) = P\{X < x\} = P_0 + P_1 + \cdots +$

جدول 1 . قانون بواسّون (تابع)

$F(x) = P(X < x)$. الترددات المتراكمة . P_x الترددات الفردية

3,5		4,0		4,5		5,0	
P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$

جدول 1 . قانون بواسّون (تابع)

الترددات للمتراكمة $F(x) = P\{X < x\}$. الترددات الفردية P_x

6,0		6,5		7,0		7,5	
P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$

جدول 1 . قانون بواسّون (تابع)

$F(x) = P \{ X < x \}$. الترددات الفردية P_x . الترددات المتراكمة

8,0		8,5		9,0		9,5		10,0	
P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$	P_x	$F(x)$
)00 3	0 0.000 3	0.000 2	0 0.000 2	0,000 1	0 0.000 1	0,000 1	0 0.000 1	ε	0 ε

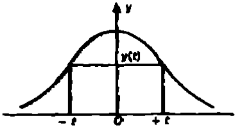
جدول 1 . قانون بواسّون (تابع)

P_x الترددات الفردية $F(x) = P(X < x)$ الترددات المتراكمة

x \ m	11 P_x	11 $F(x)$	12 P_x	12 $F(x)$	13 P_x	13 $F(x)$	14 P_x	14 $F(x)$	15 P_x	15 $F(x)$
0										
1	0,000 2	ε	0,000 1	ε						
2	0,001 0	0,000 2	0,000 4	0,000 1	0,000 2	ε	0,000 1	ε		
3	0,003 7	0,001 2	0,001 8	0,000 5	0,000 8	0,000 2	0,000 4	0,000 1	0,000 2	ε
4	0,010 2	0,004 9	0,005 3	0,002 3	0,002 7	0,001 0	0,001 3	0,000 5	0,000 7	0,000 2
5	0,022 4	0,015 1	0,012 7	0,007 6	0,007 0	0,003 7	0,003 7	0,001 8	0,001 9	0,000 9
6	0,041 1	0,037 5	0,025 5	0,020 3	0,015 2	0,010 7	0,008 7	0,005 5	0,004 8	0,002 8
7	0,064 6	0,078 6	0,043 7	0,045 8	0,028 1	0,025 9	0,017 4	0,014 2	0,010 4	0,007 6
8	0,088 8	0,143 2	0,065 5	0,089 5	0,045 7	0,054 0	0,030 4	0,031 6	0,019 4	0,018 0
9	0,108 5	0,232 0	0,087 4	0,155 0	0,066 1	0,099 7	0,047 3	0,062 0	0,032 4	0,037 4
10	0,119 4	0,340 5	0,104 8	0,242 4	0,085 9	0,165 8	0,066 3	0,109 3	0,048 6	0,069 8
11	0,119 4	0,459 9	0,114 4	0,347 2	0,101 5	0,251 7	0,084 4	0,175 6	0,066 3	0,118 4
12	0,109 4	0,579 3	0,114 4	0,461 6	0,109 9	0,353 2	0,098 4	0,260 0	0,082 9	0,184 7
13	0,092 6	0,688 7	0,105 6	0,576 0	0,109 9	0,463 1	0,106 0	0,358 4	0,095 6	0,267 6
14	0,072 8	0,781 3	0,090 5	0,681 6	0,102 1	0,573 0	0,106 0	0,464 4	0,102 4	0,362 2
15	0,053 4	0,854 1	0,072 4	0,772 1	0,088 5	0,675 1	0,098 9	0,570 4	0,102 4	0,465 6
16	0,036 7	0,907 5	0,054 3	0,844 5	0,071 9	0,763 6	0,086 6	0,669 3	0,096 0	0,568 0
17	0,023 7	0,944 2	0,038 3	0,898 8	0,055 0	0,835 5	0,071 3	0,755 9	0,084 7	0,664 0
18	0,014 5	0,967 9	0,025 5	0,937 1	0,039 7	0,890 5	0,055 4	0,827 2	0,070 6	0,748 7
19	0,008 4	0,982 4	0,016 1	0,962 6	0,027 2	0,930 2	0,040 9	0,882 6	0,055 8	0,819 3
20	0,004 6	0,990 8	0,009 7	0,978 7	0,017 7	0,957 4	0,028 6	0,923 5	0,041 8	0,875 1
21	0,002 4	0,995 4	0,005 5	0,988 4	0,010 9	0,975 1	0,019 1	0,952 1	0,029 9	0,916 9
22	0,001 2	0,997 8	0,003 0	0,993 9	0,006 5	0,986 0	0,012 1	0,971 2	0,020 4	0,946 8
23	0,000 6	0,999 0	0,001 6	0,996 9	0,003 7	0,992 5	0,007 4	0,983 3	0,013 3	0,967 2
24	0,000 3	0,999 6	0,000 8	0,998 5	0,002 0	0,996 2	0,004 3	0,990 7	0,008 3	0,980 5
25	0,000 1	0,999 9	0,000 4	0,999 3	0,001 0	0,998 2	0,002 4	0,995 0	0,005 0	0,988 8
26		1,000 0	0,000 2	0,999 7	0,000 5	0,999 2	0,001 3	0,997 4	0,002 9	0,993 8
27			0,000 1	0,999 9	0,000 2	0,999 7	0,000 7	0,998 7	0,001 6	0,996 7
28				1,000 0	0,000 1	0,999 9	0,000 3	0,999 4	0,000 9	0,998 3
29						1,000 0	0,000 2	0,999 7	0,000 4	0,999 2
30							0,000 1	0,999 9	0,000 2	0,999 6
31								1,000 0	[illegible]	0,999 8
32										

جدول 2 . كثافة احتمال قانون لابلاس - غوص

(القانون الطبيعي أو المعدل) . $y(t) = y(-t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$



t	0	1	2	3	4	5	6	7	8	9
0,	0,398 9	0,397 0	0,391 0	0,381 4	0,368 3	0,352 1	0,333 2	0,312 3	0,289 7	0,266 1
1,	0,242 0	0,217 9	0,194 2	0,171 4	0,149 7	0,129 5	0,110 9	0,094 0	0,079 0	0,065 6
2,	0,054 0	0,044 0	0,035 5	0,028 3	0,022 4	0,017 5	0,013 6	0,010 4	0,007 9	0,006 0
3,	0,004 4	0,003 3	0,002 4	0,001 7	0,001 2	0,000 9	0,000 6	0,000 4	0,000 3	0,000 2

مثلاً :

$y(1{,}3) = 0{,}171\,4$

$y(-2{,}7) = 0{,}010\,4$

جدول 2 . وظيفة توزيع قانون لابلاس - غوس



احتمال قيمة أصغر من t :

$$\pi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-t^2/2} \, dt .$$

t	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
,6	0,7257	0,7290	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9779	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

جدول قيم t الكبيرة

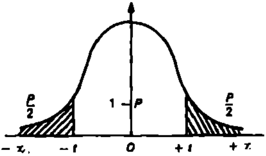
t	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,8	4,0	4,5
Π(t)	0.99865	0,99904	0,99931	0,99952	0,99966	0,99976	0,999841	0,999928	0,999968	0,999997

ملاحظة : يعطينا الجدول قيم Π(t) حيث t إيجابي . إذا كان t سلبياً يجب أخذ المتمم إلى واحد من القيمة المقروءة الجدول .

مثلاً :. pour $t = -1,37$ $\Pi(t) = 0,9147$
pour $t = 1,37$ $\Pi(t) = 0,0853$.

جدول 4 . قانون لابلاس ـ غوس

قيمة t حيث احتمال أن نتجاوز |t| هو P



P	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	∞	2,575 8	2,326 3	2,170 1	2,053 7	1,960 0	1,880 8	1,811 9	1,750 7	1,695 4
0,1	1,644 9	1,598 2	1,554 8	1,514 1	1,475 8	1,439 5	1,405 1	1,372 2	1,340 8	1,310 6
0,2	1,281 6	1,253 6	1,226 5	1,200 4	1,175 0	1,150 3	1,126 4	1,103 1	1,080 3	1,058 1
0,3	1,036 4	1,015 2	0,994 5	0,974 1	0,954 2	0,934 6	0,915 4	0,896 5	0,877 9	0,859 6
0,4	0,841 6	0,823 9	0,806 4	0,789 2	0,772 2	0,755 4	0,738 8	0,722 5	0,706 3	0,690 3
0,5	0,674 5	0,658 8	0,643 3	0,628 0	0,612 8	0,597 8	0,582 8	0,568 1	0,553 4	0,538 8
0,6	0,524 4	0,510 1	0,495 9	0,481 7	0,467 7	0,453 8	0,439 9	0,426 1	0,412 5	0,398 9
0,7	0,385 3	0,371 9	0,358 5	0,345 1	0,331 9	0,318 6	0,305 5	0,292 4	0,279 3	0,266 3
0,8	0,253 3	0,240 4	0,227 5	0,214 7	0,201 9	0,189 1	0,176 4	0,163 7	0,151 0	0,138 3
0,9	0,125 7	0,113 0	0,100 4	0,087 8	0,075 3	0,062 7	0,050 2	0,037 6	0,025 1	0,012 5

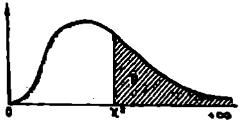
جدول قيم P صغيرة

P	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
t	3,290 5	3,890 6	4,417 2	4,891 6	5,326 7	5,730 7	6,109 4

: مثلاً : pour $P = 0,17$ $t = 1,372\ 2$.

جدول 5 . توزيع χ^2 (قانون ك. بيرسون K.Pearson)

قيمة χ^2 حيث احتمال تجاوزها هو P



v	P = 0,90	0,80	0,70	0,50	0,30	0,20	0,10	0,05	0,02	0,01
1	0,0158	0,0642	0,148	0,455	1,074	1,642	2,706	3,841	5,412	6,635
2	0,211	0,446	0,713	1,386	2,408	3,219	4,605	5,991	7,824	9,210
3	0,584	1,005	1,424	2,366	3,665	4,642	6,251	7,815	9,837	11,345
4	1,064	1,649	2,195	3,357	4,878	5,989	7,779	9,488	11,668	13,277
5	1,610	2,343	3,000	4,351	6,064	7,289	9,236	11,070	13,388	15,086
6	2,204	3,070	3,828	5,348	7,231	8,558	10,645	12,592	15,033	16,812
7	2,833	3,822	4,671	6,346	8,383	9,803	12,017	14,067	16,662	18,475
8	3,490	4,594	5,527	7,344	9,524	11,030	13,362	15,507	18,168	20,090
9	4,168	5,380	6,393	8,343	10,656	12,242	14,684	16,919	19,679	21,666
10	4,865	6,179	7,267	9,342	11,781	13,442	15,987	18,307	21,161	23,209
11	5,578	6,989	8,148	10,341	12,899	14,631	17,275	19,675	22,618	24,725
12	6,304	7,807	9,034	11,340	14,011	15,812	18,549	21,026	24,054	26,217
13	7,042	8,634	9,926	12,340	15,119	16,985	19,812	22,362	25,472	27,688
14	7,790	9,467	10,821	13,339	16,222	18,151	21,064	23,685	26,873	29,141
15	8,547	10,307	11,721	14,339	17,322	19,311	22,307	24,996	28,259	30,578
16	9,312	11,152	12,624	15,338	18,418	20,465	23,542	26,296	29,633	32,000
17	10,085	12,002	13,531	16,338	19,511	21,615	24,769	27,587	30,995	33,409
18	10,865	12,857	14,440	17,338	20,601	22,760	25,989	28,869	32,346	34,805
19	11,651	13,716	15,352	18,338	21,689	23,900	27,204	30,144	33,687	36,191
20	12,443	14,578	16,266	19,337	22,775	25,038	28,412	31,410	35,020	37,566
21	13,240	15,445	17,182	20,337	23,858	26,171	29,615	32,671	36,343	38,932
22	14,041	16,314	18,101	21,337	24,939	27,301	30,813	33,924	37,659	40,289
23	14,848	17,187	19,021	22,337	26,018	28,429	32,007	35,172	38,968	41,638
24	15,659	18,062	19,943	23,337	27,096	29,553	33,196	36,415	40,270	42,980
25	16,473	18,940	20,867	24,337	28,172	30,675	34,382	37,652	41,566	44,314
26	17,292	19,820	21,792	25,336	29,246	31,795	35,563	38,885	42,856	45,642
27	18,114	20,703	22,719	26,336	30,319	32,912	36,741	40,113	44,140	46,963
28	18,939	21,588	23,647	27,336	31,391	34,027	37,916	41,337	45,419	48,278
29	19,768	22,475	24,577	28,336	32,461	35,139	39,087	42,557	46,693	49,588
30	20,599	23,364	25,508	29,336	33,530	36,250	40,256	43,773	47,962	50,892

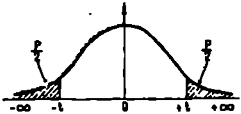
ملاحظة : v هو عدد درجات الحرّية .

إذا كان v محصوراً بين 30 و100 ، نقرّ بأنّ $\sqrt{2\chi^2} - \sqrt{2v-1}$ يتوزّع تقريبا حسب القانون الطبيعي للمركز للمختصر $(m = 0, \sigma = 1)$.

إذا كان v أكبر من 100 ، نقرّ بأنّ $(\chi^2 - v) / \sqrt{2v}$ يتوزّع تقريباً حسب القانون الطبيعي للمركز للمختصر $(\sigma = 1, m = 0)$.

جدول 6 . توزيع ستودنت ـ فيشر

قيمة t حيث احتمال أن تتجاوز |t| هو P



n	P=0,90	0,80	0,70	0,60	0,50	0.40	0.30	0.20	0.10	0.05	0.02	0.01
1	0.158	0.325	0.510	0.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	0.142	0.289	0.445	0.617	0.816	1.061	1.386	1.886	2.920	4.303	6.965	9.925
3	0.137	0.277	0.424	0.584	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841
4	0.134	0.271	0.414	0.569	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604
5	0.132	0.267	0.408	0.559	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032
6	0.131	0.265	0.404	0.553	0.718	0.906	1.134	1.440	1.943	2.447	3.143	3.707
7	0.130	0.263	0.402	0.549	0.711	0.896	1.119	1.415	1.895	2.365	2.998	3.499
8	0.130	0.262	0.399	0.546	0.706	0.889	1.108	1.397	1.860	2.306	2.896	3.355
9	0.129	0.261	0.398	0.543	0.703	0.883	1.100	1.383	1.833	2.262	2.821	3.250
10	0.129	0.260	0.397	0.542	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169
11	0.129	0.260	0.396	0.540	0.697	0.876	1.088	1.363	1.796	2.201	2.718	3.106
12	0.128	0.259	0.395	0.539	0.695	0.873	1.083	1.356	1.782	2.179	2.681	3.055
13	0.128	0.259	0.394	0.538	0.694	0.870	1.079	1.350	1.771	2.160	2.650	3.012
14	0.128	0.258	0.393	0.537	0.692	0.868	1.076	1.345	1.761	2.145	2.624	2.977
15	0.128	0.258	0.393	0.536	0.691	0.866	1.074	1.341	1.753	2.131	2.602	2.947
16	0.128	0.258	0.392	0.535	0.690	0.865	1.071	1.337	1.746	2.120	2.583	2.921
17	0.128	0.257	0.392	0.534	0.689	0.863	1.069	1.333	1.740	2.110	2.567	2.898
18	0.127	0.257	0.392	0.534	0.688	0.862	1.067	1.330	1.734	2.101	2.552	2.878
19	0.127	0.257	0.391	0.533	0.688	0.861	1.066	1.328	1.729	2.093	2.539	2.861
20	0.127	0.257	0.391	0.533	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845
21	0.127	0.257	0.391	0.532	0.686	0.859	1.063	1.323	1.721	2.080	2.518	2.831
22	0.127	0.256	0.390	0.532	0.686	0.858	1.061	1.321	1.717	2.074	2.508	2.819
23	0.127	0.256	0.390	0.532	0.685	0.858	1.060	1.319	1.714	2.069	2.500	2.807
24	0.127	0.256	0.390	0.531	0.685	0.857	1.059	1.318	1.711	2.064	2.492	2.797
25	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.316	1.708	2.060	2.485	2.787
26	0.127	0.256	0.390	0.531	0.684	0.856	1.058	1.315	1.706	2.056	2.479	2.779
27	0.127	0.256	0.389	0.531	0.684	0.855	1.057	1.314	1.703	2.052	2.473	2.771
28	0.127	0.256	0.389	0.530	0.683	0.855	1.056	1.313	1.701	2.048	2.467	2.763
29	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.311	1.699	2.045	2.462	2.756
30	0.127	0.256	0.389	0.530	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750
∞	0.12566	0.25335	0.38532	0.52440	0.67449	0.84162	1.03643	1.28155	1.64485	1.95996	2.32634	2.57582

ملاحظة . n هو عدد درجات الحرية .

جدول 7 . أعداد الصدفة (1)

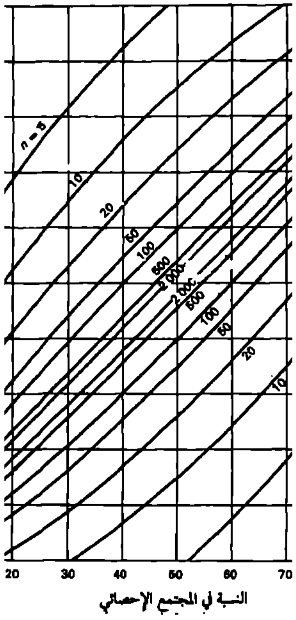
Trente-cinquième mille								
5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
85 19	48 74	55 24	89 69	15 53	00 20	88 48	95 08	00 47
34 51	40 44	62 93	65 99	72 64	09 34	01 13	09 74	90 65
96 79	38 24	77 00	70 91	47 43	43 82	71 67	49 90	37 09
81 85	50 47	36 50	91 19	09 15	98 75	60 58	33 15	51 44
80 04	21 49	54 91	77 85	00 45	68 23	12 94	23 44	36 88
52 73	06 41	37 47	47 31	52 99	89 82	22 81	86 55	99 09
52 72	49 11	30 93	33 29	54 17	54 48	47 42	04 79	18 64
64 07	85 32	05 96	54 79	57 43	96 97	30 72	12 19	41 70
92 29	71 11	64 10	42 23	23 67	01 19	20 58	35 93	39 46
32 91	95 28	42 36	98 59	66 32	15 51	46 63	57 10	83 55
04 62	24 87	44 85	45 68	41 66	19 17	13 09	63 37	15 33
55 88	25 01	15 77	12 90	69 34	36 93	52 39	36 23	99 73
18 93	86 98	99 04	75 28	30 05	12 09	57 35	90 15	98 07
35 47	16 32	20 16	78 52	82 37	26 33	67 42	11 93	35 61
82 18	06 61	54 67	03 66	76 82	90 31	71 90	39 27	97 85
58 65	27 70	93 57	59 00	63 56	18 79	85 52	21 03	03 16
89 23	76 46	97 70	00 62	15 35	97 42	47 54	60 60	78 12
65 62	81 29	69 71	95 53	53 69	20 95	66 60	50 70	22 97
98 15	05 64	43 32	74 03	44 63	52 38	67 59	56 69	11 14
41 48	64 79	62 26	87 86	94 30	43 54	26 98	61 38	63 44
02 24	67 85	88 10	34 01	54 53	23 77	33 11	19 68	13 50
87 56	19 19	19 43	70 25	24 29	48 22	44 81	35 40	33 23
25 10	87 27	77 28	05 90	73 03	95 46	88 82	25 02	05 00
14 03	17 80	47 85	94 49	89 55	10 27	19 50	20 37	02 71
93 40	45 43	04 57	17 03	34 54	83 91	69 02	90 72	98 45

الألف الـ 36								
5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
04 66	68 52	70 11	97 01	55 36	63 49	42 68	82 15	48 64
98 82	61 64	40 50	42 48	96 84	82 42	55 15	72 34	90 96
93 55	89 63	47 92	88 42	00 08	21 52	27 28	77 48	02 42
97 55	27 91	15 20	96 25	48 75	49 95	88 68	36 09	66 17
55 98	33 02	36 99	11 84	07 71	40 65	95 54	01 90	14 32
16 48	38 14	94 74	00 37	24 88	26 40	05 87	01 87	00 82
28 29	14 36	09 42	22 63	85 40	79 23	60 18	58 89	60 95
67 37	40 50	74 11	57 07	54 90	55 60	75 66	74 59	43 34
38 38	35 63	71 92	51 61	07 57	33 15	47 80	14 72	67 27
47 95	21 81	99 54	84 68	49 46	04 87	23 10	93 18	34 62
12 18	78 69	61 17	41 02	82 98	57 15	80 65	08 18	25 81
67 03	44 53	15 36	14 27	47 96	35 38	29 07	84 99	51 14
58 34	23 47	96 09	36 91	82 76	68 90	21 61	55 66	74 17
02 16	80 53	35 89	06 64	54 32	96 97	74 19	33 04	06 70
70 60	97 25	37 17	72 52	39 87	15 15	98 30	51 57	06 42
16 18	55 02	72 66	63 80	21 24	20 23	18 13	84 73	83 73
86 60	42 36	12 67	10 64	97 65	96 18	41 67	59 91	42 75
35 52	20 78	72 37	23 78	53 42	92 51	26 14	61 35	49 00
53 45	09 38	08 72	03 92	86 92	91 44	96 12	68 34	30 86
16 25	24 40	90 62	85 78	10 68	26 14	78 07	47 97	94 91
93 74	37 82	93 68	50 32	56 81	15 70	78 54	37 33	97 30
08 59	17 46	26 25	32 70	13 62	73 02	34 58	46 18	89 59
05 77	58 49	14 59	77 89	35 73	54 07	30 65	59 68	82 98
48 94	94 27	76 81	68 16	97 85	03 80	49 25	10 37	43 88
57 45	47 95	42 13	86 48	02 36	50 36	36 32	85 38	04 15

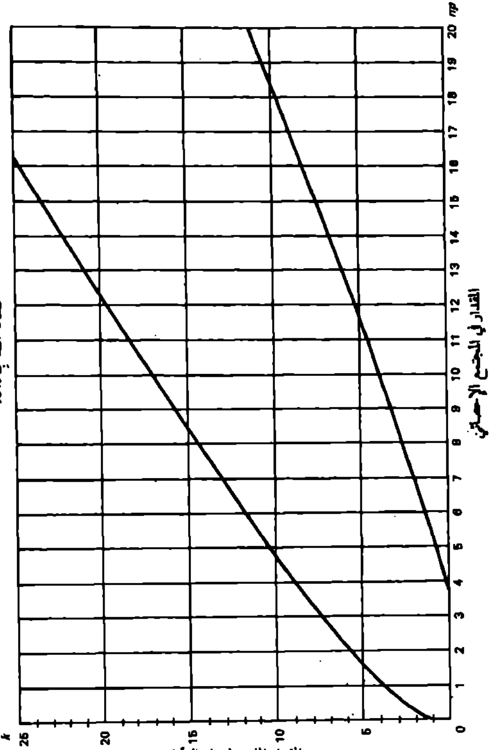
(1) مقتطف من

...puters, edited by E. S. Pearson, D. Sc. No. XXIV, *Tables of Random Sampling* ... M. G. Kendall et B. Babington Smith, Cambridge University Press, 1946.

اللوحة البيانية رقم 1 . تقدير النسبة p
لدرجة الثقة بـ 95%



اللوحة البيانية رقم 2 . تقدير النسبة p (تقريب من قانون بواسون)
لسعة الثقة ب 95%



بيبليوغرافيا موجزة

مؤلفات عامّة

G. CALOT, *Cours de statistique descriptive*, Paris, Dunod, 1973.

G. CALOT, *Cours de calcul des probabilités*, Paris, Dunod, 1971.

H. CRAMER, *Mathematical methods of statistics*, Princeton University Press, 1961.

C. FOURGEAUD et A. FUCHS, *Statistique*, Paris, Dunod, 1967.

C. FOURGEAUD et G. HANSEL, *Statistique*, licence ès sciences économiques 2e année, Paris, Librairie Dey, 1969.

C. FOURGEAUD et P. LECOINTE, *Statistique*, licence ès sciences économiques 3e année. Paris, Librairie Dey, 1970.

H. GUITTON, *Statistique*, Paris, Dalloz, 1971.

M. G. KENDALL and A. STUART, *The advanced theory of statistics*, London, Ch. Griffin, 2 vol., 1961, 1963.

W. L. L'ESPERANCE, *Modern statistics for business and economics*, New York, Macmillan Co., 1971.

W. MASIERI, *Notions essentielles de statistique et calcul des probabilités*, Paris, Sirey 1973.

W. C. MERRIL and K. A. FOX, *Introduction to economic statistics*, New York, John Wiley and Sons, 1970.

A. M. MOOD and F. A. GRAYBILL, *Introduction to the theory of statistics*, New York, McGraw-Hill, 1963.

E. MORICE et F. CHARTIER, *Méthode statistique*, 2 vol., Paris, Imprimerie nationale, 1954.

J. MOTHES, *Prévisions et décisions statistiques dans l'entreprise*, Paris, Dunod, 1962.

P. ROSENSTIEHL et J. MOTHES, *Mathématiques de l'action*, Paris, Dunod, 1968

R. SCHLAIFER, *Probability and statistics for business decisions*, New York, McGraw Hill, 1959.

S. S. WILKS, *Elementary statistical analysis*, Princeton University Press, 1961.

S. S. WILKS, *Mathematical statistics*, New York, John Wiley and Sons, 1962.

G. U. YULE and M. G. KENDALL, *An introduction to the theory of statistics*, London Ch. Griffin, 1945.

الأبحاث الإحصائية . . الفصلان V وVII

W. G. COCHRAN, *Sampling techniques*, New York, John Wiley and Sons, 1963.

W. E. DEMING, *Sampling design in business research*, New York, John Wiley and Sons, 1960.

J. DESABIE, *Théorie et pratique des sondages*, Paris, Dunod, 1971.

M. H. HANSEN, W. HURWITZ and W. G. MADOW, *Sample survey methods and theory*, New York, John Wiley and Sons, 1953.
Volume I. *Methods and applications*. Volume II. *Theory*.

L. KISH, *Survey sampling*, New York, John Wiley and Sons, 1965.

L. L. VANCE and J. NETER, *Statistical sampling for auditors and accountants*, New York, John Wiley and Sons, 1961.

فهرست

الموضوع **الصفحة**

هذا الكتاب

ما يميّز هذا الكتاب هو أنـه يقدّم ،
ضمن إطار عملي وموجّـه نحو التطبيق ، فكرة شاملة عن
مختلف مظاهر التفكير الإحصائي ، وهو بهذا يساعد على تسهيل
مهمّـة الطالب والإحصائي بالحدّ من عدد المصادر المتنوّعة التي يضطران
للرجوع إليها .

كما أنّـه خلال عرضه للتطبيقات العملية ، لا يتمسّـك كثيراً بالأداة الرياضية المعقّـدة التي تنفر القارىء وتضجره دون أن تكون ضرورية لفهم سيرورة التفكير ووضعه موضع التطبيق . ومن هنا فهو يطمح إلى شرح « التقنيات الإحصائية » تحت شكلها العملي ودون رجوع مبالغ فيه إلى الأداة الرياضية ، هذه التقنيات التي أضحت معرفتها اليوم ضرورية للمسؤولين والموظّـفين في أكثر من مجال إداري واقتصادي .

إنّـه إذن يقدّم وسائل التحليل الإحصائي منطلقاً في عرضه للطرق الإحصائية من حساب الاحتمالات وقوانين التوزيعات مروراً بالبحوث الإحصائية وطرق تطبيقها ووصولاً إلى تحليل السلاسل الزمنية بالاستناد إلى الإنحدار والإرتباط الإحصائيين .

كلّ هذا نجده مرفقاً بأمثلة عديدة
ومتنوّعة معالجة بتفاصيلها بغية إعطاء القارىء صورة ملموسة عن
أفكار المؤلّـف ودليلاً واضحاً من أجل التطبيق
على حالات من الواقع .

الكِتاب للنشر والتوزيع
المؤسسة الجامعية للدراسات والنشر والتوزيع